

# Large-Scale Scientific endeavours: the production and dissemination of advance computer sciences knowledge

---

**Arturo Sánchez Pineda**

*Università di Udine, INFN and ICTP, Italy*

*E-mail: [arturos@cern.ch](mailto:arturos@cern.ch)*

Perform big data analysis and visualisation on your own computer? Yes, you can! Commodity computers are now very powerful in comparison to only a few years ago. On top of that, the performance of today's software and data development techniques facilitates complex computation with fewer resources. Cloud computing is not always the solution, and reliability or even privacy is regularly a concern. While the Infrastructure as a Service (IaaS) and Software as a Service (SaaS) philosophies are a key part of current scientific endeavours, there is a misleading feeling that we need to have remote computers to do any kind of data analysis. One of the aims of the ATLAS Open Data project is to provide resources — data, software and documents — that can be stored and executed in computers with minimal or non-internet access, and in as many different operating systems as possible. This approach is viewed as complementary to the IaaS/SaaS approach, where local university, students and trainers' resources can be used in an effective and reproducible way — making the High Energy Physics and Computer Sciences fields accessible to more people. We present the latest developments in the production and use of local Virtual Machines and Docker Containers for the development of physics data analysis. We also discuss example software and Jupyter notebooks, which are in constant development for use in classrooms, and students and teachers computers around the world. Also, we showcase how those tools are developed and in use by international efforts (IPPOG, ICTP-PWF and CEVALE2VE.org) in Latin America and Europe to bring knowledge in countries like Mexico, Venezuela, Colombia, Peru, Ecuador, Argentina and Uruguay with support of the ICTP, CERN and many universities in the region.

*Artificial Intelligence for Science, Industry and Society, AISIS2019*

*October 21-25, 2019*

*Universidad Nacional Autónoma de México, Mexico City, México*

## 1. Introduction

The members of the ATLAS Experiment [1, 2], at the Large Hadron Collider (LHC) [3] of the European Organisation for Nuclear Research (CERN) [4], have developed an extensive and always-evolving outreach programme [5], that –with the collaboration of CERN and other partners– is dedicated to disseminating the physics, the techniques, the results and the implications of the science and engineering that its members produce every day.

The outreach activities performed by the ATLAS Collaboration, and many other multinational scientific endeavours are relevant for several reasons. Some of them are moral because resources, funding and support come from the taxpayers of the members countries of the collaboration that directly or indirectly sponsor the programme over 25 years [5]. Another of the crucial reasons is more pragmatic and equally relevant: to ensure continuity of the science over the decades; engaging and training present and future qualified and motivated human resources all over the world. At the same time, ensuring that the knowledge transfer is happening to as many people as possible, and doing so, raising the probabilities that such knowledge can be applied effectively in other areas of science, development and culture.

The above statement sounds ambitious and it is. This is why concrete plans and projects have been expanding to perform such tasks as effectively as possible. One of them has been the release of datasets, educational — web-based and offline — tools and documentation under the so-called ATLAS Open Data project [6, 7, 8, 9, 10].

Another example of successful outreach project is the use of video-conferencing technology to “bring” the ATLAS Control Room, the ATLAS detector, its physicists and engineers to the classroom in a real-time conversation between them and primary, high school, university students and their teachers. This project is called the ATLAS Virtual Visit [11].

Nowadays, both projects work together to create an educational resource that every year is used by thousands of students, who –remotely– have a close view of what we do as researchers. is the use of video-conferencing technology to “bring” the ATLAS Control Room, the ATLAS detector, its physicists and engineers to the classroom in a real-time conversation between them and primary, high school, university students and their teachers.

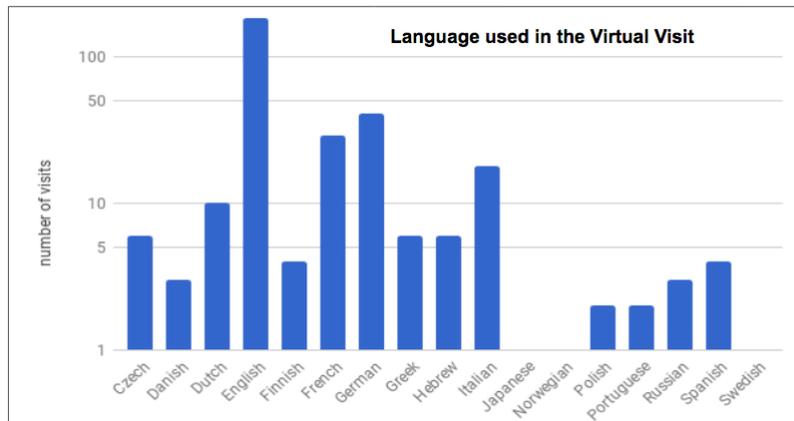
We will give a look at how the merging of different technologies, philosophies and ideas allow building successful STEM projects worldwide.

## 2. ATLAS Virtual Visits

The ATLAS Virtual Visits program [11] is a simple but powerful idea: to establish a video conference between ATLAS researchers based at CERN and a university classroom, a school or a public event anywhere in the world with an internet connection. The Virtual Visit (VV) is an online conversation, held in participants’ preferred language, between those physicists from ATLAS and students and the general public ready to listen and to ask questions (see Figure 1). An ATLAS VV provides access to the ATLAS Control Room or the ATLAS experimental cavern. During this one-hour conversation, the researchers at CERN show the ATLAS Control Room (on the surface) or the ATLAS Detector (100 meters underground) to the participants and explain the physics, technology, operations and challenges of the present experiment and future upgrades (see Figure 2).



**Figure 1:** Several events in Latin America (the flags represent some of the countries where educational events were helped in 2017 or 2018). On those, the ATLAS Virtual Visits have been used to complement other events like public seminars, IPPOG [13] Masterclasses [14] and training workshops for students.

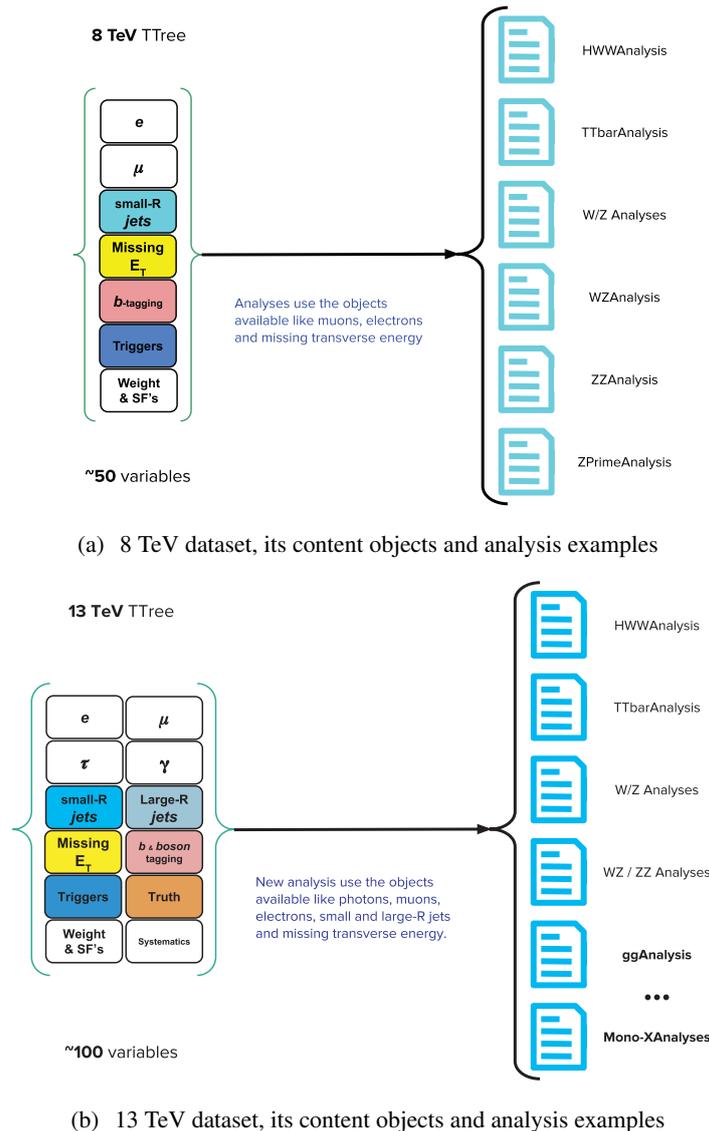


**Figure 2:** On top: a plot with the languages used to address the participants in ATLAS VV during 2018. On the bottom, two different views of the hosts in an ATLAS VV: underground next to the ATLAS Detector (bottom-left) and at the surface in front of the ATLAS Control Room (bottom-right).

POS(AISISS2019)040

### 3. ATLAS Open Data project

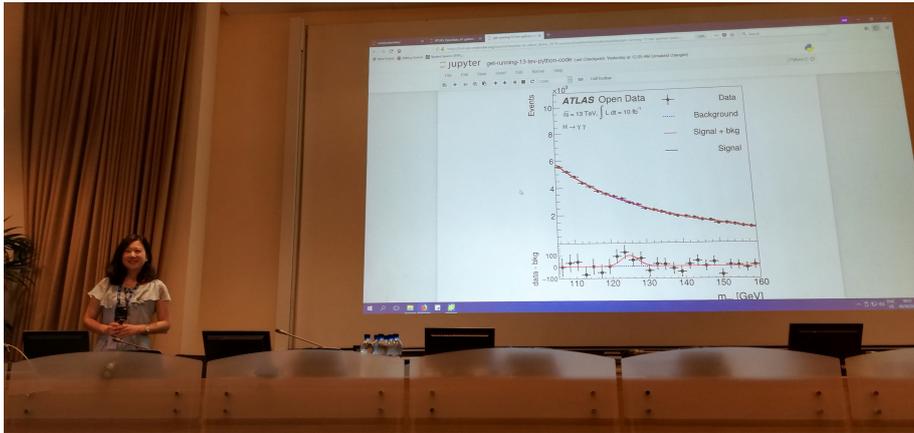
The ATLAS Open Data project[6, 8, 9, 12] is an effort of the ATLAS Collaboration dedicated to designing a holistic educational programme in HEP for undergraduate and master students all over the world. This effort involves the design, production, testing, validation and analysis of real and simulated Monte Carlo (MC) datasets that will be released to the public following the ATLAS Policy on Open Access with focus on educational and training objectives [16, 15]. Tools and examples and web-based applications are produced and delivered alongside the datasets.



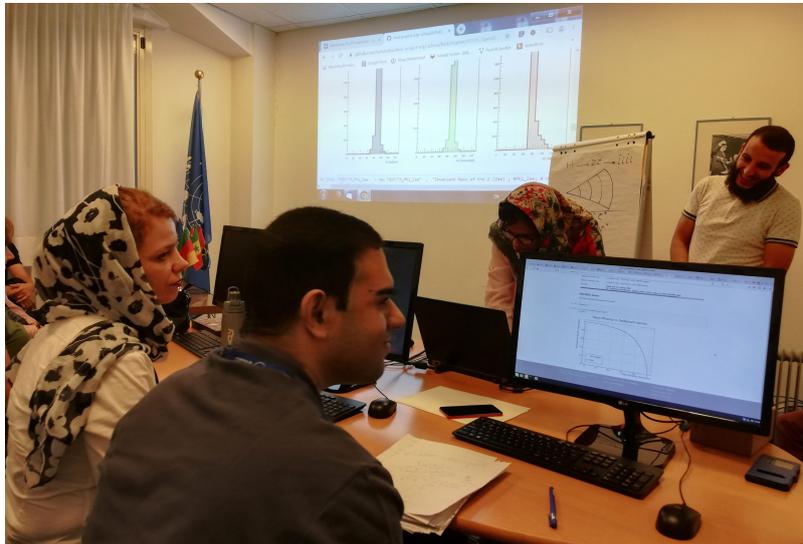
**Figure 3:** Dataset and analysis examples for the 8 TeV and 13 TeV of the ATLAS Open Data project.

The dataset is designed, tested and constructed to maximise the number of physics processes and analysis examples that can be studied and replicated in a classroom (see Figure 3), while keeping them as small — in terms of digital size — as possible.

This way they can fit on and be processed by a commodity computer; as used in labs and by individual students and professors in institutions around the globe. Together with those datasets, software applications, manuals and documentation are produced in ways that can be used with and without constant access to the Internet. In this way, we ensure that, even in not so ideal conditions, students and teachers (see Figure 4) can profit from the dataset and the tools to learn and explore HEP and the relevant computer science. Up to now, ATLAS has released  $1 \text{ fb}^{-1}$  of real data and several more  $\text{fb}^{-1}$ 's of MC datasets at a centre-of-mass energy of 8 TeV[6]. And more recently  $10 \text{ fb}^{-1}$  of real data and hundreds of MC datasets at a centre-of-mass energy of 13 TeV[12], together with an analysis frameworks, Jupyter [19] notebooks, dedicated Virtual Machines (VM) and simple web-based and desktop applications in its website [9] (see Figures 5 and 6).



(a) Final ATLAS Open Data Summer Students' report presentation at CERN, August 2019 [17].



(b) Presentation of exercises of ICTP workshop's students using ATLAS Open Data resources. Trieste, August 2019 [18].

**Figure 4:** The pictures above show two cases where students and trainers are using and developing ATLAS Open Data educational resources in HEP. In both cases it is possible to see the use of data visualisation web-based tools to (a) demonstrate the final result of a particle physics analysis and (b) discuss of the potential of the effect of a particular selection on different MC datasets, including multivariate analysis techniques.

## 4. HEP Outreach collaborations

In the next sections, some outreach and educational programmes are briefly described. They have been developed independently and currently work in partnerships to enhance their specific activities or to create new ones. It is relevant also to mention that most of these initiatives are on a voluntary basis. So, their members donate their time and expertise to the enhancement of education in places far from their home institutions. In many of the cases, the students, researchers and professors involved as instructor or developer of the resources are using this kind of initiatives to give back to the countries of origin and helping the global community of future researchers. In other cases, the outreach program is supported by the collaboration, and the approach is more centred (but not limited) to members institutions. Of course, since the resources are Open Access, they can be used by anybody else with the will to learn from them.

### 4.1 IPPOG International Masterclasses

The ALICE, ATLAS, CMS and LHCb experiments participate actively in the IPPOG [13] International Masterclasses [14]. Each year more than 14000 high school students in about 50 countries come to one of about 220 nearby universities or research centres for one day to unravel the mysteries of particle physics in a simple, practical and engaging session of data analysis. Lectures from active scientists give insight in topics and methods of basic research on the nature of matter and forces, enabling the students to perform measurements on real data from particle physics experiments themselves. At the end of each day, like in international research collaboration, the participants join a video conference for discussion and combination of their results, moderated by physicists in research centres in the Americas and Europe.

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

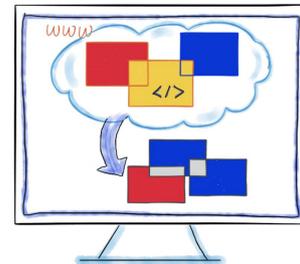
- Data samples in ROOT n-tuple format
- Software and Jupyter [Notebooks](#) in Python and C++ to analyse the samples and produce physics analysis
- JavaScript (JS) applications to produce cut-and-count analysis
- Virtual Machines with several Linux-based OS and ROOT CERN analysis framework
- [GitHub](#)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

#### Data & Tools Repository

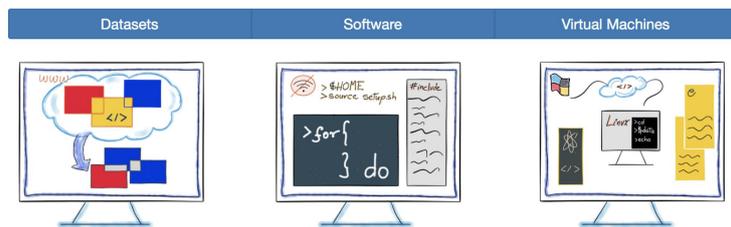
Here you have in a single place all the necessary pieces in order to start you physics analysis in a more complete way. Look into the data like an ATLAS particle physicist!

In this section, you can find where to download:

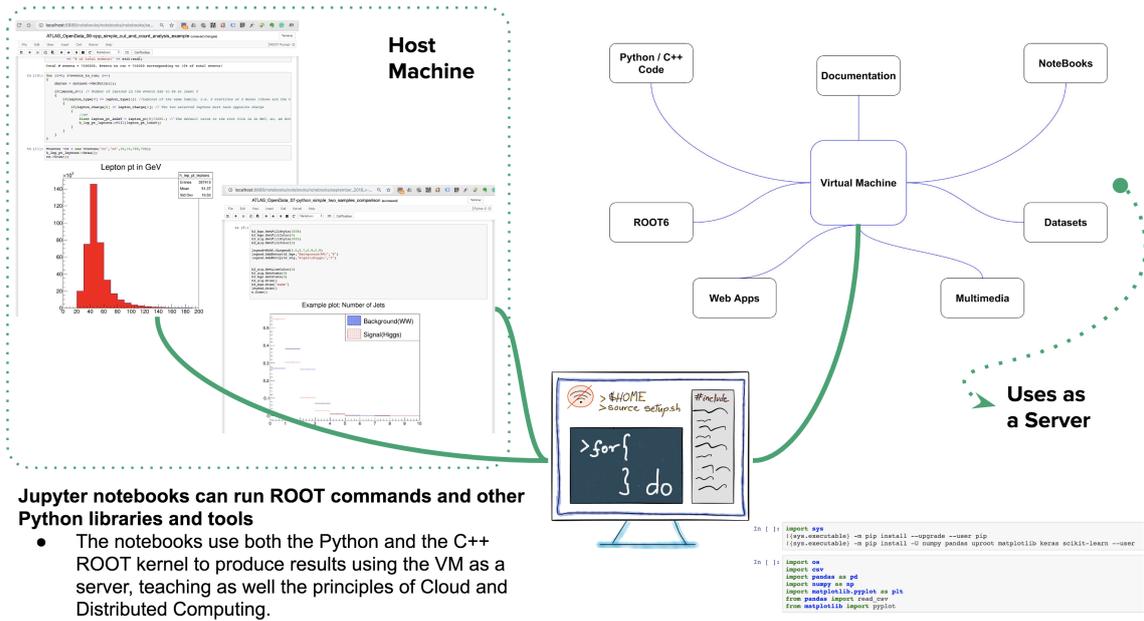
- The complete collection of available datasets
- The different analysis software
- The virtual machines to perform physics searches



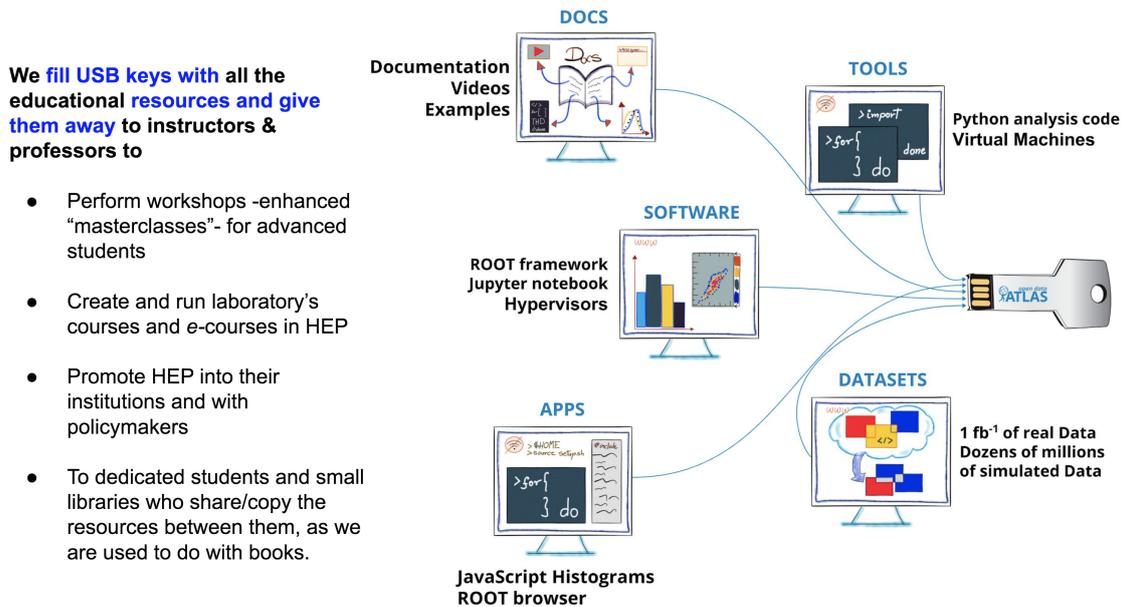
#### Downloads



**Figure 5:** A summary of the datasets and educational resources delivered by the ATLAS Open Data project. Where the aim is to allow the users to start to explore and analyse the data with the minimum setup possible, also, the documentation shows examples and multimedia material to begin to execute the software and tools.



(a) A schematic view on how the Virtual Machine allows the user to have her Jupyter server to work offline without complex setup in the computer. It is possible to use more than one kernel, ROOT and no-ROOT based.



(b) The Open Key allows the portability of those resources to institutions and individuals with limited Internet access—also, the possibility to have a borrowing system for the resources. Similar to books in a library, students can borrow the pen drive to take the datasets and tools to their computers at home. As well, it allows us to set up a standard computer lab for a workshop in a matter of minutes.

**Figure 6:** The pictures above show how the ATLAS Open Data resources, and other Open Access tools, can be used by the students and instructors in their institutions. Also, the portability and flexibility to set individual and groups working spaces to read the datasets and execute the resources, even under no-Internet conditions.

POS(AISIS2019)040

## 4.2 CEVALE2VE

The Centro de Altos Estudios de Altas Energías (CEVALE2VE) [20] is a virtual research and learning community created to support the new generation of Venezuelan researchers in HEP. It also attempts to contribute to the knowledge transfer of fundamental physics and the regional modernisation of university education with the promotion of scientific communities in Latin America. The project organisation involves several academic institutions in Venezuela, Colombia, Peru, Guatemala, Ecuador and Mexico, to reach a broad regional audience. Although geographically scattered in different academic institutions of Europe and North America, CEVALE2VE integrates a group of Venezuelan researchers currently involved in projects related to the LHC. Its main goal is to stimulate and widen regional physics postgraduate education and research, the first step of which has been the implementation of the online university course “Introduction to Particle Physics“ [20]. The main objective is to develop a virtual research and learning community to promote education and research in the fields of HEP and computer sciences in the Latin American scientific community [21]. The CEVALE2VE activities can be summarised as: Education and interest in research: positively influence and stimulate physics students’ interest in HEP research.

- **Awareness of opportunities:** educate and invigorate the students with further study opportunities in physics and possible career paths in research.
- **Create networks and collaborations:** between the European and North American institutions and research centres with their partner universities in Latin America.
- **Modernisation of education and training:** through the use and development of a web platform with e-learning open access tools.
- **Policymaking:** working towards formalising the involvement of Latin American institutions in HEP experiments.

## 4.3 ICTP Physics Without Frontiers

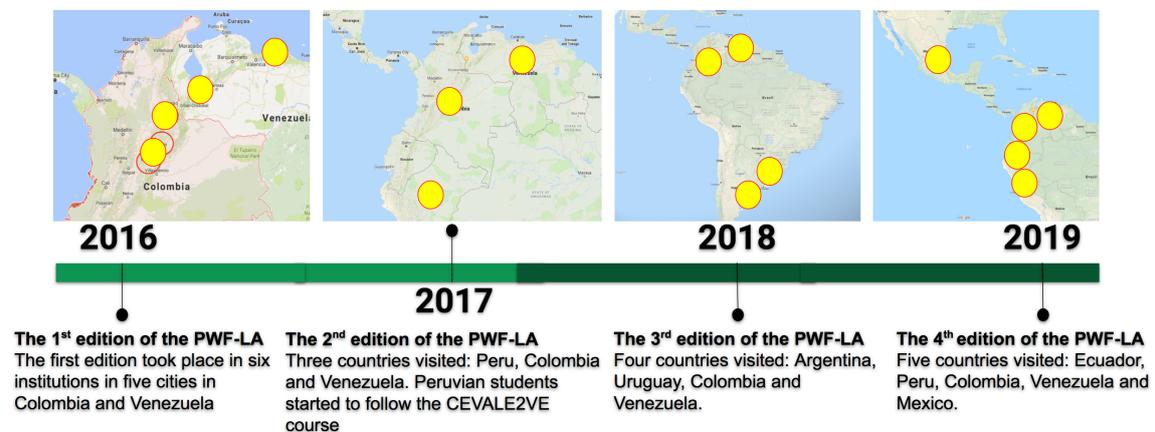
The ICTP [22] Physics Without Frontiers (PWF) [23] programme is an outreach set of activities designed by different researchers, especially in the developing world, under the guidance and supervision of ICTP coordinators. This is a programme that designates resources and support to approved applications coming from an inclusive community of researchers, professors and science communicators, among others. The PWF programme works to inspire, train and motivate physics and mathematics university students worldwide, to help build the next generation of scientists. Each project is unique, developed with the country’s specific needs in mind.

PWF organises projects working with volunteer scientists, who are PhD students, postdoctoral researchers, or lecturers from all over the world passionate about promoting physics and mathematics. PWF has worked with over 5000 students worldwide in 17 different countries [23].

### 4.3.1 Physics Without Frontiers projects

- **University Courses:** PWF partners with physics and mathematics departments that are unable to teach specific courses due to a lack of resources. The programme arranges an expert to visit the university to teach the course and where possible train local researchers and teachers.
- **Schools:** PWF organise pedagogical schools at universities for local undergraduate and master students across the country. The programme aims to expose and train the students in research fields they lack access to.
- **Roadshows:** Roadshows are organised by groups of young researchers, passionate about advocating their research field back in their home [24]. The roadshow centres around intensive one-day activities in universities across the host country, in addition to outreach and public events (see Figure 7).

The PWF programme has the final aim of breaking down barriers in accessing physics and mathematics research. This means connecting with all students and scientists worldwide, and with those from less represented groups in physics and mathematics within countries, due to factors such as gender, economic background, ethnicity, or those who may lack access due to geographic barriers or isolation within a region.



**Figure 7:** A summary of the ICTP PWF and CEVALE2VE programme in Latin America. The set of maps shows the evolution of the roadshows in multiple cities in the continent [24].

## 5. Summary

Promoting and performing HEP's educational programmes, training and outreach in a multi-cultural and globalised environment is a constant challenge, and a source of opportunities to innovate and to collaborate between individuals, institutions, programmes and countries to take the best of their efforts and join them, creating useful, broad and maintainable projects in the long term.

The ATLAS Outreach group has joined resources and efforts to create activities that involve people from many educational, cultural, geographical and institutional backgrounds, improving the knowledge transfer and the exchange of ideas and the development of human resources into the HEP sector. Other partnerships have been established with external outreach groups, making the impact of such programmes even larger. Nowadays, outreach communities like the IPPOG, CEVALE2VE, ICTP PWF and many individuals are executing successful projects that invite even more researchers and institutions to get involved [21, 25, 26, 27].

This positive evolution in open access resources, activities and events is increasingly getting the attention and support of the policymakers, and large scientific collaborations that are evident in dedicated sessions in important conferences like the one that this article is intended to document. We have been dedicated a large and ever-increasing amount of time and effort to develop computing solutions for those programmes, and they are now running in hundreds of computers all over the planet. That year by year is going in an exponential way.

The overall observation is that the different groups have been finding ways to enhance the education of diverse communities, thanks to the collaboration and constant feedback among them. No surprise there, this is how science moves forward also.

## References

- [1] ATLAS Collaboration, *The ATLAS Experiment at the CERN LHC*. 2008 *JINST* **3** S08003.
- [2] *The ATLAS Collaboration at LHC*. <https://atlas.cern>
- [3] *The Large Hadron Collider (LHC)*. <https://home.cern/science/accelerators/large-hadron-collider>
- [4] *The European Organisation for Nuclear Research (CERN)*. <https://home.cern>
- [5] The ATLAS Collaboration, *ATLAS: A 25-Year Insider Story of the LHC Experiment Chapter 11: ATLAS Collaboration: Life and its Place in Society* Advanced Series on Directions in High Energy Physics: Volume 30, 2019. [https://doi.org/10.1142/9789813271807\\_0011](https://doi.org/10.1142/9789813271807_0011)
- [6] ATLAS Collaboration, Review of ATLAS Open Data Dataset. *ATL-OREACH-PUB-2016-001*.
- [7] ATLAS Collaboration, Review of ATLAS Open Data 8 TeV datasets, tools and activities. *ATL-OREACH-PUB-2018-001*
- [8] Article: *Explore LHC data on new ATLAS educational platform*. July, 2016. <http://cern.ch/go/MSN7>
- [9] *The ATLAS Open Data web portal*. <http://opendata.atlas.cern>
- [10] *The ATLAS Open Data release at 13 TeV*. ATLAS Experiment releases 13 TeV Open Data for Science Education. Highest-energy particle-collision data ever released through open access.
- [11] *The ATLAS Virtual Visits project*. *Nucl.Part.Phys.Proc.* 273-275 (2016) 2578-2580
- [12] ATLAS Collaboration, Review of the 13 TeV ATLAS Open Data release. *ATL-OREACH-PUB-2020-001*.
- [13] *The International Particle Physics Outreach Group (IPPOG) web portal*. <http://ippog.org>
- [14] *The IPPOG Masterclasses web portal*. <https://physicsmasterclasses.org>

- [15] CERN Courier, *Preserving the Legacy of Particle Physics*. Page 33. March-April 2019. <https://cerncourier.com/wp-content/uploads/2019/07/CERNCourier2019MarApr-digitaledition.pdf>
- [16] The ATLAS Data Policy. <http://opendata.cern.ch/record/413/files/ATLAS-Data-Policy.pdf>
- [17] *2019 CERN Summer Student for the ATLAS Open Data group's final presentation*. <https://cds.cern.ch/record/2685132>
- [18] *The 2019 ICTP "The CODATA-RDA Research Data Science Advanced Workshops on Bio-informatics, Climate Data Sciences, Extreme Sources of Data and Internet of Things(IoT)/Big-Data Analytics"*. <http://indico.ictp.it/event/8847>
- [19] *The Jupyter Open Source project*. <https://jupyter.org/>
- [20] A. R. Sanchez Pineda, The CEVALE2VE case. ATL-OREACH-PROC-2017-001
- [21] Article: *Boosting high-energy physics education around the world with ATLAS Open Data*. July, 2018. <http://cern.ch/go/7nv8>
- [22] *the Abdus Salam International Centre for Theoretical Physics (ICTP) web portal*. <https://www.ictp.it/>
- [23] *The ICTP Physics Without Frontiers (PWF) web portal*. <https://www.ictp.it/physics-without-frontiers.aspx>
- [24] Article: *Physics Outreach: Colombia and the ICTP Physics Without Frontiers + CEVALE2VE*. November, 2018. <http://cern.ch/go/St6p>
- [25] M. R. Di Domenico Franco, Reconstruction of the invariant masses of bosons of the Standard Model using public data from ATLAS Open Data. CERN-THESIS-2017-239
- [26] I. Garcia, 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 8 TeV. CERN-THESIS-2017-217
- [27] M. O. Evans, Enabling Open Science with the ATLAS Open Data project at CERN. CERN-THESIS-2018-099