PROCEEDINGS OF SCIENCE

ATLAS Open Data – a genuinely collaborative approach for the creation of educational resources

Meirin Oan Evans[a,*] **on behalf of the ATLAS Collaboration**

[a]*University of Sussex,*
Falmer, Brighton, Great Britain
E-mail: meirin.oan.evans@cern.ch

The ATLAS Open Data project aims to deliver open-access resources for education and outreach in High Energy Physics and related computer sciences. Because the resources – data, software and documentation – target students and instructors, they must be tested by students and instructors before being released. One of the most effective production ways we have found is to promote on-site and remote training schemes such as high-school work experience and summer schools, university projects and PhDs qualification tasks. Those programmes and alliances allow the construction of a platform that relies on the expertise of ATLAS members and the invaluable contribution of students that help to test and build resources that hundreds of their peers use around the world. We present how multiple training programmes inside and outside CERN helped and continue to help create the ATLAS Open Data project and the lessons learnt so far on how to continue implementing these kinds of programmes.

The European Physical Society Conference on High Energy Physics (EPS_HEP2021)
26-30 July 2021
Online conference, jointly organized by Universität Hamburg and the research center DESY

[*]Speaker

PoS(EPS-HEP2021)899

1. Motivation

These proceedings outline how students and the ATLAS [1] Open Data group have come together to co-create educational resources (Figure 1), by taking a tour around the world of different students who have contributed to this effort. These proceedings focus on co-creation for the first experimental data from Run 2 of the LHC [2].

Figure 1: The ATLAS Open Data team has come together with students from around the world to create educational resources.

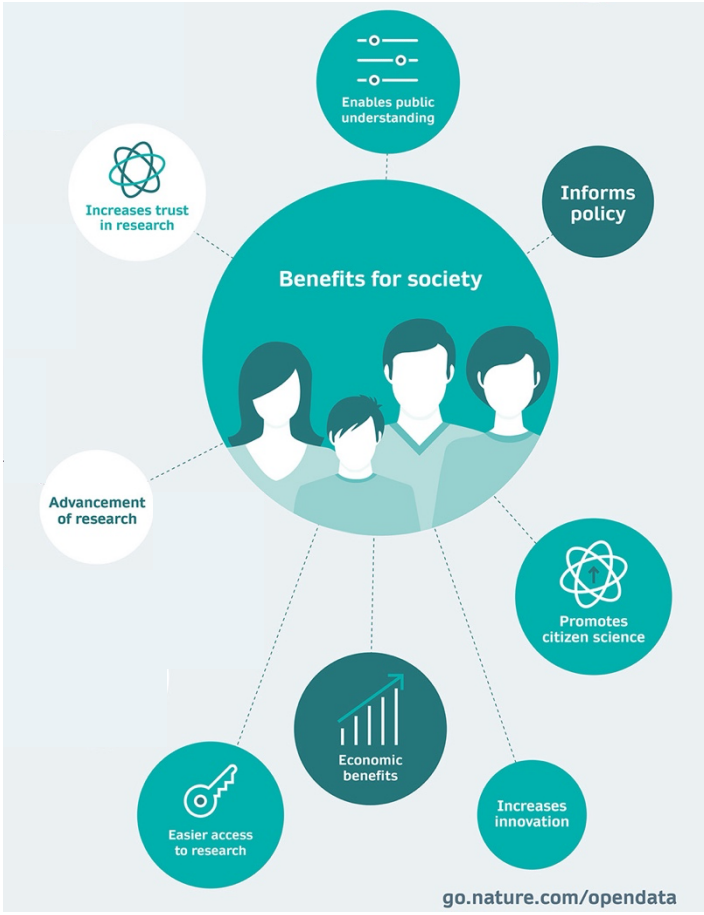

Figure 2: Benefits for society of Open Data [3].

Open Data are a crucial cornerstone of science and their benefits to society are numerous (Figure 2). They enable public understanding and inform policy. They promote citizen science and increase innovation. They provide economic benefits and easier access to research, even advancement of research, and increase trust in research and scientists.

Co-creation is about creating resources *with* and *for* students. The educational resources co-created with students include the ATLAS Open Data website [4], ATLAS Open Data 13 TeV documentation [5], ATLAS Open Data GitLab [6] and ATLAS Open Data GitHub [7]. These resources are highlighted in Figure 3.

PoS(EPS-HEP2021)899

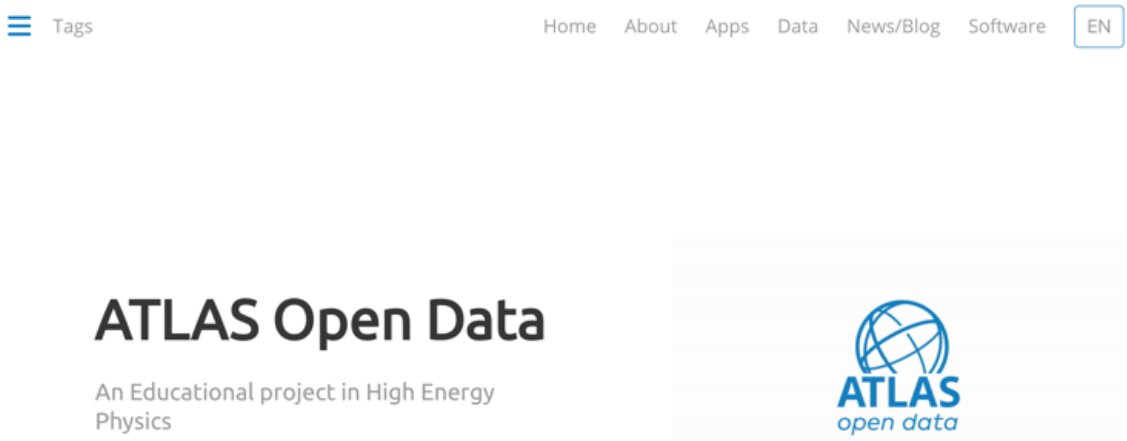

Figure 3: The ATLAS Open Data website [4] collects different educational resources co-created with students.

2. Co-creation examples

The goal of this conference talk was to take the audience around the world in 15 minutes, showcasing different examples of co-creating ATLAS Open Data educational resources. The map of students involved in this co-creation is shown in Figure 4.

Figure 4: Students co-creating ATLAS Open Data educational resources come from all over the world.

The first step to co-create 13 TeV ATLAS Open Data was to produce the Open Data files, including real and simulated data. Two first-year PhD students, Meirin and Even, from Britain and Norway respectively, were responsible for this task. This included producing files for 10 fb^{-1} of real data, along with approximately 120 Monte Carlo simulated physics processes, split into seven collections [8]. Each file gave access to approximately 90 variables. This wealth of files

and variables allows analyses of H→WW, ttbar, W, Z, WZ, ZZ, diphoton, SUSY and exotic events, as well as many more.

Having produced Open Data files, a code was needed to analyse them. A CERN summer student Ya-Feng, from Taiwan, produced seven ROOT analysis examples to accompany the 13 TeV ATLAS Open Data release [9]. The goal of these examples was to educate and train others.

Even with data files and analysis examples, further tools are useful for the release of Open Data. Two more CERN summer students, Shodruz and Yixin, from Tajikstan and China respectively, developed an ATLAS Open Data Jupyter notebook interface with ROOT [10]. These tools teach histogram plotting, analysis methods, code explanation and more.

Alongside data and tools, documentation is paramount for Open Data. Ten 16–17-year-old students, from BHASVIC college in Britain, helped incorporate student voice in the 13 TeV ATLAS Open Data documentation. They first did this by including questions in the text, so that a future student following the documentation could test their understanding. Their next step was to embed curriculum links in the text, so that students and teachers could tie their learning from ATLAS Open Data back into their curricula. To finish, they peer reviewed each other's improvements to the documentation, by providing suggestions and comments. Various documentation pages were co-created in this way [11].

Documentation made more accessible to younger students can only be a positive thing. During an internship before starting university, Iago from Brazil conceived of a visual way to explain some tables in the 13 TeV ATLAS Open Data documentation. He created Sankey diagrams of Higgs boson decay modes to explain which decay modes are being chased in one of the ATLAS Open Data analyses. These diagrams were then incorporated into the documentation [12].

With complete documentation, it is important to test it with the target audience. Three work experience students Ander, Kip and William from Britain did exactly this. In doing so, they enhanced the documentation by incorporating any suggestions they had. Having tested and improved documentation, more students can be reached by translating the documentation. Two work experience students from Switzerland, Brieuc and Maud, were responsible for translating the documentation into French [13].

Now that ATLAS Open Data are ready to be used by students, the next step is multi-user deployment. Another 1st year PhD student, Giovanni from Italy, is developing an infrastructure to deploy reproducible educational data-analysis platforms at small or medium projects and institutions. This means downloading or streaming ATLAS Open Data resources to an external cloud, and then into institutional resources. More information can be found in Ref. [14].

ATLAS Open Data educational resources would not be the same without the help of numerous students from the Americas. Over various years, the resources were stress-tested, particularly in locations where high-speed internet is not commonplace, therefore cloud computing is *not* a solution for everybody. There were too many students to name, but they were

based all over Argentina, Colombia, Ecuador, Mexico, Peru, Uruguay and Venezuela. Further details of this programme can be found in Ref. [15]. Some photographs from this project are shown in Figure 5.

Figure 5: Some photographs from the project stress-testing ATLAS Open Data educational resources with students across the Americas [15].

Students discussed throughout this section are shown in Figure 6.

Figure 6: Students co-creating ATLAS Open Data educational resources come from all over the world.

3. Summary

There are a number of key challenges to consider whilst co-creating educational resources with students:

- How to ensure our learning resources are accessible without guidance from physicists?
- How to incorporate our resources into more schools?
- How can we emphasise the **teaching of more than just particle physics** – skills in computing, programming, coding, analysis, data science, machine learning and more?
- How can we empower students with **additional skills** – such as presenting, poster design, teamwork, research and more?
- How can we reach more students that need it, such as from under-represented groups?

The ATLAS Open Data project gives access to cutting-edge research data for use in education, along with educational resources to analyse the data. Co-creation of these resources with students leads to: 1) better resources, 2) better experience for students, 3) a fun experience for all involved. Can more of us co-create resources with students?

References

[1] ATLAS Collaboration: *JINST* **3** (2008) S08003.

[2] Large Hadron Collider: https://home.cern/science/accelerators/large-hadron-collider.

[3] Research Data Policies: https://www.springernature.com/gp/authors/research-data-policy.

[4] ATLAS Open Data Website: http://opendata.atlas.cern/.

[5] ATLAS Open Data 13 TeV Documentation: http://opendata.atlas.cern/release/2020/documentation/.

[6] ATLAS Outreach data and tools GitLab: https://gitlab.cern.ch/atlas-outreach-data-tools.

[7] ATLAS Outreach data and tools GitHub: https://github.com/atlas-outreach-data-tools.

[8] ATLAS Open Data, 13 TeV Open Datasets http://opendata.atlas.cern/release/2020/documentation/datasets/files.html.

[9] ATLAS Outreach C++ framework 13 TeV: https://github.com/atlas-outreach-data-tools/atlas-outreach-cpp-framework-13tev.

[10] MyBinder notebooks collection open data: https://mybinder.org/v2/gh/atlas-outreach-data-tools/notebooks-collection-opendata/master.

[11] Data Visualisation: http://opendata.atlas.cern/release/2020/documentation/visualization.

[12] The Higgs Boson: http://opendata.atlas.cern/release/2020/documentation/visualization/the-higgs-boson.html.

[13] ATLAS Open Data Website Français: https://atlas-open-data-docs.github.io/WoW/fr/.

[14] A proposal for Open Access data and tools multi-user deployment using ATLAS Open Data for Education: https://indico.cern.ch/event/948465/contributions/4323945/.

[15] South American Wide HEP programme: ictp.it/physics-without-frontiers/current-country-projects/south-american-wide-pwf-hep-programme.