LHC Open Data for the world to see

Meirin Oan Evans[a,*]**, on behalf of the ALICE, ATLAS, CMS and LHCb Collaborations**

[a]*School of Mathematical and Physical Sciences, University of Sussex,*
Falmer, Brighton, Great Britain

E-mail: meirin.oan.evans@cern.ch

Open Data are a crucial cornerstone of science and their benefits to society are numerous. For education and outreach with LHC Open Data, we focus on benefits such as easier access to research, enabling of public understanding and promotion of citizen science. This contribution will build up examples of using LHC Open Data for education, training and outreach, from advanced high school students, to PhD students and interested publics outside formal education.

The Eighth Annual Conference on Large Hadron Collider Physics -LHPC2020
25-30 May, 2020
online

---

*Speaker

1. The importance of Open Data

Open Data are a crucial part of science and their benefits for society are numerous (Figure 1). Open Data 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. For education and outreach, we focus on benefits such as enabling public understanding, promoting citizen science and easier access to research.
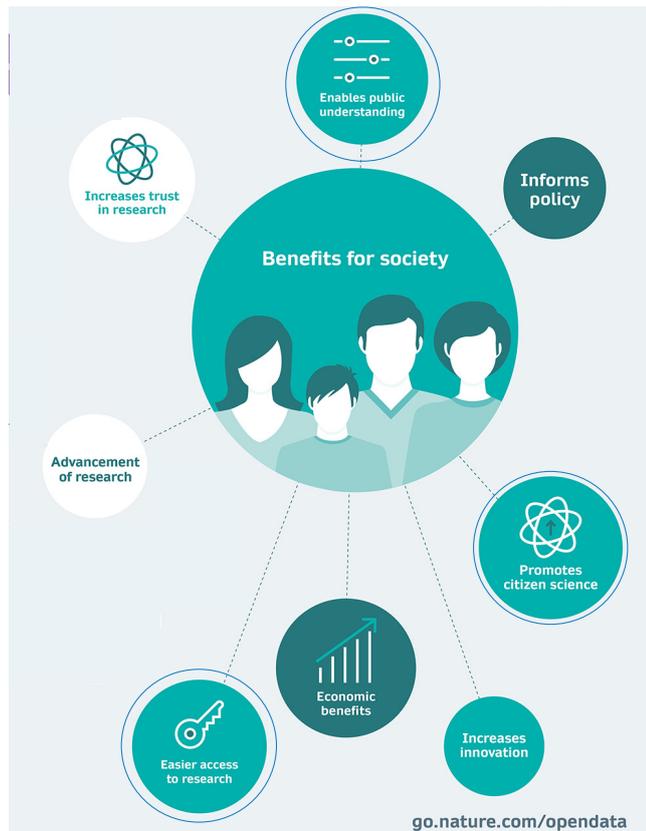
Figure 1: Benefits for society of Open Data.

2. Educational motivation for LHC Open Data

As well as general educational motivations for Open Data, there are specific educational motivations for LHC Open Data, summarised in Figure 2. Our target audience ranges from advanced high school students to PhD students. Activities with Open Data can vary from a few hours to multiple
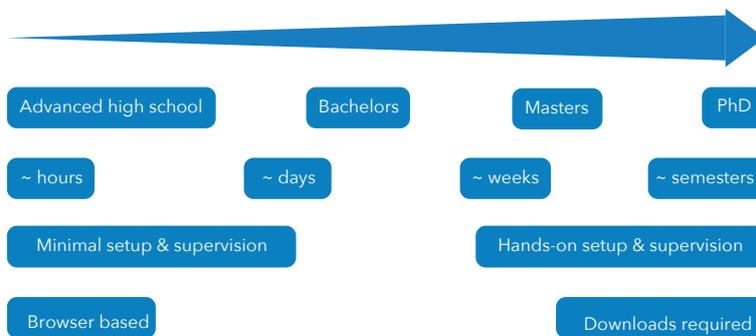
Figure 2: Our motivation is to reach a broad audience, from high school to PhD students.

semesters. Some activities only require minimal setup and supervision, whereas others need more hands-on setup and supervision. The most accessible activities are browser based, whereas if students want to interact with all available data, downloads are required.

The story of LHC Open Data starts with International Masterclasses [1]. These are days where high school students from around the world get the chance to be particle physicists for a day. They do a particle-physics analysis and then have a video conference with other universities that have hosted students anywhere in the world and some moderators at CERN. The question for LHC

Open Data teams has been ``how can students go beyond masterclasses in terms of time spent, difficulty, physics level, amount of data and research independence?"

The answer is the CERN Open Data Portal. It is a web portal [2], and it is the place to get started. Anyone coming to this website can choose to explore data, software, computing environments and documentation. Visitors can choose to focus on particular experiments or on topics such as data science.

3. Educational examples of LHC Open Data
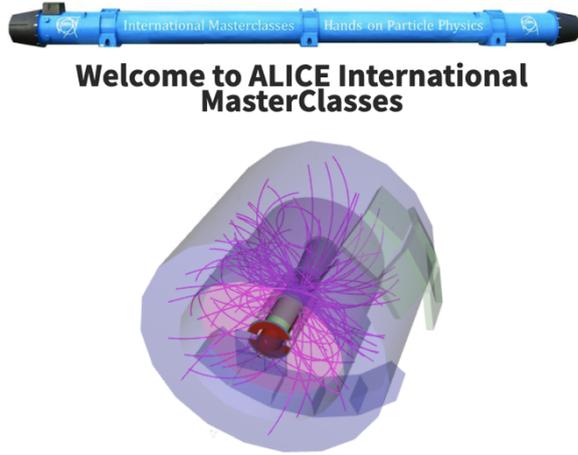
Figure 3: Screenshot of the ALICE International Masterclass to search for strange particles [3].
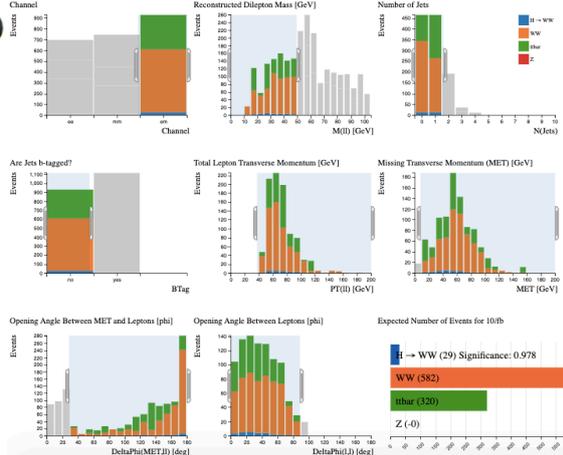
Figure 4: Screenshot of the ATLAS Histogram Analyser to search for the Higgs boson with only your mouse [4].
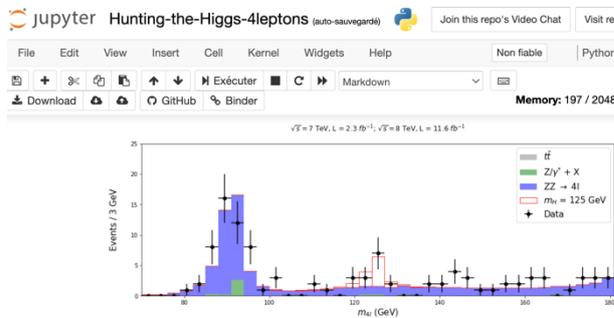
Figure 5: Screenshot of CMS Open Data that allow students to see a peak around 125 GeV build within minutes [6].
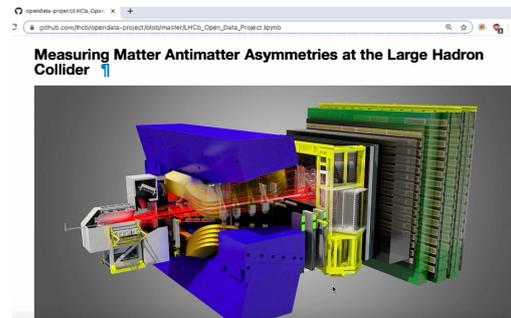
Figure 6: Screenshot of an analysis to measure matter-antimatter asymmetries at LHCb [7].

The first educational example of LHC Open Data is a masterclass searching for strange particles in ALICE [3] (Figure 3). As with other masterclasses, it is aimed at high school students, but could be used as an exercise for undergraduates. Software and data are packaged together into one download. Together they are a tool for teachers. Of course, support from physicists helps!

Moving from high school to university, we present the ATLAS Histogram Analyser [4] (Figure 4). This is a web tool for fast, cut-based data analysis. Students visualise data using online histograms, allowing them to search for the Higgs boson using only their mouse!

Another example is CMS Open Data education [5], open-source exercises written in multiple languages. Current languages include English, Finnish, Spanish, German and Greek. Data are in simplified formats such as CSV files. Students can see the peak around 125 GeV in Figure 5 build within minutes of opening a webpage [6]!

Another example is an analysis looking for matter-antimatter differences (Figure 6). The goal is to analyse B-meson decays. A Jupyter [7] notebook provides a guided analysis of LHCb data [8]. This could be used, for example, in a third-year lab course.

4. Educational motivation for LHC Open Data

Many challenges lie ahead for using LHC Open Data in education. For example:

- How can we ensure our tools are accessible without guidance from physicists?
- How can we incorporate our resources into more university (and pre-university!) courses?
- How can we spread our tools and resources into wider use?
- How can we teach more than physics – computing, data science, machine learning...?

5. Your turn to get started!

Open Data are a crucial part of science, especially science education and outreach. LHC Open Data tackle the education aspect by providing resources for a range of students, from high school to PhD. Students need only navigate to opendata.cern.ch to get started. However, the community using LHC Open Data for education has many challenges to address going forward. Even with these challenges, LHC Open Data are ready for educators to use in their teaching today.

References

[1] International Particle Physics Outreach Group, 2020, International Masterclasses, Physics Masterclasses, https://physicsmasterclasses.org/ (Accessed on 1 September 2020).

[2] CERN, 2020, CERN Open Data, http://opendata.cern.ch/ (Accessed on 1 September 2020).

[3] ALICE Collaboration, 2020, ALICE Masterclass, Welcome to ALICE International MasterClasses, https://alice-masterclass.web.cern.ch/ (Accessed on 1 September 2020).

[4] ATLAS Collaboration, 2020, Histogram Analyser, ATLAS Open Data 13 TeV Documentation, http://opendata.atlas.cern/release/2020/documentation/visualization (Accessed on 1 September 2020).

[5] CMS Collaboration, 2020, CMS Open Data Education, https://github.com/cms-opendata-education (Accessed on 1 September 2020).

[6] CMS Collaboration, 2018, Hunting the Higgs 4 leptons, Higgs Hunting – an example of scientific research, https://mybinder.org/v2/gh/cms-opendata-education/cms-jupyter-materialsenglish/master?filepath=Exercises-with-open-data/Advanced (Accessed on 1 September 2020).

[7] Project Jupyter, 2020, https://jupyter.org (Accessed on 1 September, 2020).

[8] LHCb Collaboration, 2017, Measuring Matter Antimatter Asymmetries at the Large Hadron Collider, LHCb Open Data Project, https://github.com/lhcb/opendata-project (Accessed on 1 September 2020).