

## In Pursuit of Authenticity – CMS Open Data in Education

---

**Peitsa Veteli** <sup>a,\*</sup> and **Kati Lassila-Perini** <sup>a</sup>

<sup>a</sup>*Helsinki Institute of Physics,*

*Gustaf Hällströmin katu 2, 00014 Helsinki, Finland*

*E-mail:* [peitsa.veteli@helsinki.fi](mailto:peitsa.veteli@helsinki.fi), [kati.lassila-perini@cern.ch](mailto:kati.lassila-perini@cern.ch)

There are some universally acknowledged problems in school sciences. In developed countries worldwide, young people are not interested in studying STEM subjects. Whether that is because of perceived lack of personal relevance, disconnect from the actual fields of study, “sanitized” school practices or other factors is a matter of debate, but it is eminently clear that, as educators, we must do our best to combat this trend. In this paper, we discuss how open data from the CMS experiment has been used in education and present feedback from Finnish teachers who have received training in using these freely available programming resources to bring modern physics into their teaching. The main focus here is on the teachers' perception of authenticity in the use of “real world” research data, although there is an additional benefit of learning general scientific methods and cross-disciplinary data handling skills as well.

Keywords: open data, education, authenticity, CMS, Jupyter Notebooks, teacher training, secondary school

*40th International Conference on High Energy physics - ICHEP2020*  
July 28 - August 6, 2020  
Prague, Czech Republic (virtual meeting)

---

\*Speaker

## **1. Introduction**

Students' motivation to learn STEM subjects and pursue careers related to STEM has been in global decline for a long time, especially in developed countries [1,2]. While the exact reasons can be debated, there is a long list of potential causes, such as too controlling a school environment, perceived difficulty of the scientific subjects combined with general anxiety, a notion of them being irrelevant for their lives, a lack of relatable work goals or a lack of a sense of meaning in the activities or methods employed. For decades now, education research literature has advocated the use of methods that stimulate and help maintain student interest through feelings of relevance and authenticity. These terms are, however, often vaguely defined and thrown around rather casually to mean something that is done "like in the real world" in contrast to more contrived examples of "school science". In this paper, we present a case for utilizing openly available particle physics data as a source of authentic inquiry work with meaningful programming, materials created for this purpose and present feedback from teachers that have received training in their use.

## **2. On authenticity in science education**

A recent review of the science education research literature [3] found out that the use of authenticity as a term can be separated into distinct categories, with the three most commonly used meanings with respect to authentic activities being the following:

- comparable to methods and challenges of practicing scientists
- grounded in the world of the students
- involving inquiry-based methods.

The first category especially has traditionally been promoted, since what purpose does science education serve if it is not even comparable to the science it purports to represent? This question naturally oversimplifies the realities of school life and the constant tug of war between the demands of academia, curricula, industries and society at large as to what, how and at what pace should be learned during the formative years. Tensions exist between the first two categories. One plausible reason for decreasing motivation among students to participate in science education may be its traditional reliance on methods that work for the comparatively few students actually intending to pursue academic careers later on, but which do not engage the rest of the student body effectively at all [4]. The remaining students, who will later become politicians, bureaucrats, entrepreneurs and so forth, might not necessarily need to do as much "sciencing" themselves in their chosen careers. Would it still not benefit us all as a society if they were more positively interested in the practices of those who do pursue careers in science? Appreciation of science by the general public is a crucial element in being a data-literate citizen in the 21<sup>st</sup> century, one that even extends beyond the simple benefit of noticeably greater interest in the science subjects during school years if it allows them to participate in science education with ways they find intrinsic, personal meaning in as well as offering a sense of "realness" in relation to the world around them.

### 3. CMS Open Data in education

Since 2014, the CMS experiment at CERN has published unprecedented amounts of high-quality, research-grade data under open licensing for anyone to use [5]. For educational purposes, especially with secondary students, the raw data is dauntingly complex to use, but it can and has been also published in simplified formats that can be accessed without extensive skills in programming or high energy physics. These simplified datasets are still large and messy enough to preserve their authentic qualities for inexperienced users, allowing for emergent educational challenges comparable in spirit to those faced by career scientists. With these resources at hand, the Helsinki Institute of Physics' Education and Open Data project has since been creating similarly freely accessible educational materials that draw on the expertise of both science educators and particle physics researchers to provide teachers more tools to bring a touch of modern, cutting-edge research to their classrooms. Materials can be found at the Github repositories: <https://github.com/cms-opendata-education>

The materials combine meaningful programming for a purpose with the often abstract subject of particle physics. By utilising datasets that are too large to go through by hand, students are prompted to analyze them in a manner not uncommon to particle physicists. The guiding principles in our approach to material creation and interactions with the teachers, which arise from experience in the field as well as existing literature, are as follows:

- **Openness is the default.** Our materials are made to showcase, suggest and inspire, so that teachers can make of them what they want in a way that suits their particular style and preferred topics.
- **Transversal skills are at the heart of data education.** While students might initially only try out exercises related to particle physics (e.g. spotting Higgs bosons from background events or testing the energy spectra of various particles), they will find that similar data analysis skills will serve them quite well in other subjects as well. Cognitive studies have shown that humans tend to associate skills with specific topics rather than easily adapting what they know to other contexts [6], but such a practice can be ameliorated through engaging and motivating exercises applied to different contexts. Such skills are globally gaining traction as a recognized curriculum objective in other aspects of digital literacy as well.
- **General is better than proprietary.** We do not want to produce simple demonstrations or require specific coding skills that are only applicable to our materials, even where this might be "the correct way" from a technical standpoint. Especially in the secondary school context, there are many ongoing changes related to digitalisation that are a burden for students and teachers alike; thus we must try to support them in this transitional period by offering tools that have value beyond the lessons where they have been employed. To this end, our programming is done with Python instead of domain-specific software such as ROOT [7], both because it is a widely used language in

science and because it is easy to expand on for whatever purposes might suit the students' needs.

- **Teachers have too much on their plate already.** Initial use of the materials must not require lengthy installations or pose other technical barriers to entry beyond having a laptop and a working internet connection. From experience, we know that teachers do not want to spend too much time getting into a new tool set, doubly so into a whole new skill set like programming, unless they can immediately see some value and ease of use in it. Therefore, we use Jupyter Notebooks [8] as a relatively light, yet versatile, environment to quickly run Python on the fly, with free browser-based virtual workspaces like MyBinder [9] and Google Colab [10] doing the work. With prepared materials, only a shareable hyperlink is needed and the students can be working on their own in a couple of minutes.

### 3.1. Teacher trainings

Getting people to adopt new tools and approaches often requires a hands-on effort from the original advocates. To this end, we have organized several in-service teacher training courses in Finland as well as participated in the High-School Teacher Programmes at CERN between 2016 and 2020 as a fairly efficient way to promote these ideas: if you can inspire a teacher to adopt a new method, they can influence hundreds of students in turn. Here, we highlight the Finnish cases and how responses gathered therein relate to the earlier notion of authenticity as a desirable attribute in science education.

The trainings usually last two full working days with 10-20 participants guided by 2-3 workshop leaders, who can be working scientists, PhD students or otherwise involved experts. Most secondary level teachers are not researchers or programmers themselves, so the first day tends to prove quite taxing, with many participants browsing through the existing materials and trying out the data for themselves. On the second day, after the “shock and horror” phase has abated, there is more variation between those who are seeking a deeper understanding of what is provided and those who are experimenting with creating their own exercises. These events are also excellent opportunities for the science teachers to exchange ideas and discuss future collaborations. Given how many of these early participants often come from schools that are part of the national Finnish CERN network and already have an interest in such projects, such a co-operative spirit is not surprising.

### 3.2. Teacher responses

Of the approximately 70 participants thus far, 32 gave feedback relevant for this paper. The answers were gathered via an e-form at the end of the event without guiding them towards a particular topic, letting the teachers raise various themes of interest. In general, the feedback has been universally positive, with the participants reporting that they have gained valuable ideas from the course. A more detailed analysis of these answers is currently in the making.

- Of the 32 teachers who provided feedback, only 20% felt fluent in programming before the training, yet more than 50% noted that programming was a part of their students' education. The state of things on that front is still somewhat in flux, as Finland instated programming as a mandatory part of the high-school curriculum only in 2016, but how it is being implemented varies from school to school.
- Only 16% felt that their school activities correlated well with the “real world”. Another 16% questioned if such a relation was even necessary at all, while most of the participants felt a definite detachment between the classroom science and the “real world”.
- After the workshops, 66% felt that they would like to use programming tools like these with real data, while one-third were still on the fence. It is noteworthy, though, that those who expressed more hesitancy were of that mindset more because of time constraints and the challenge of finding a good place for it in their practice, not because of the methods themselves. Only one answer out of the 32 was entirely in the negative.

Typical sample answers (translated from Finnish) are as follows:

- “I knew it [open data] existed, but hadn't thought that it could be used in teaching...”
- “It's very good that it's real data instead of something artificial and made up like you so often see in math and physics problems. A challenge lies in learning the basics first, maybe, with those made up examples... or should they?”
- “Using real world data would give the students a realistic picture of what it means to do science and research.”
- “Science comes closer to the student. Their confidence rises. A sense of 'I too can do this!' 'I can understand this!' can come to the fore if you can present the topics well.”
- “Using open data in teaching particle physics helps the students understand the challenges and possibilities of modern physics.”
- “I feel the way these exercises are made is more important than the content of the data. If you can analyse it scientifically, it advances their understanding of doing science in general.”

In future research, a follow up is planned to see if this sentiment and training translates into school practice.

#### 4. Summary

As a whole, the ideas brought forth by the teachers align well with the notion of authenticity as used by the science education research community. Coming to terms with hands-on exercises for a subject like particle physics was deemed valuable, as was the act of having to work like scientists with real, complex data and conducting occasional Google searches to make sense of how to program further. While these views are of course filtered through the perspective of teachers who are experts in their own craft, their feedback aligns reasonably well with the limited results we have seen from student workshops and internship programmes,

where the students have been quite excited by the opportunity to do something entirely different from their normal science studies. In particular, we would like to highlight the observation of empowering a sense of discovery among girls, who might not otherwise have engaged in hobbies or developed other extracurricular habits allowing them to come into much contact with the possibilities of programming. Further studies on the benefits of student-centered open data activities in fostering scientific curiosity in such cases would be interesting. When connecting the wider field of educational research with open data, it is also worth mentioning how well the answers gathered herein align with Atenas, Havemann and Priego's ideas on fuelling personal growth as critical citizens [11] by fostering critical thinking, increasing data analysis skills, understanding the research methods and nature of scientific work, and interlinking them with societal growth to better enable democratic practices that create equal opportunity for all.

## 5. Acknowledgements

The authors would like to thank the Finnish National Agency for Education for their support and encouragement through the years.

## References

- [1] J.M. Hellgren & S. Lindberg (2017), *Motivating students with authentic science experiences: changes in motivation for school science*, Research in Science & Technological Education, 35:4
- [2] S. Sjöberg & C. Schreiner (2010), *The ROSE project. An overview and key findings*
- [3] J. Anker-Hansen & M. Andréé (2019), *In Pursuit of Authenticity in Science Education*, Nordic Studies in Science Education, 15
- [4] G. Aikenhead (2005), *Research Into STS Science Education*, Educación Química, 16
- [5] CERN Open Data portal, <https://opendata.cern.ch>
- [6] J.L. Docktor & J.P. Mestre (2014), *Synthesis of discipline-based education research in physics*, Phys. Rev. ST Phys. Educ. Res 10, 020119
- [7] ROOT Data Analysis Framework <https://root.cern/>
- [8] Project Jupyter <https://jupyter.org/>
- [9] MyBinder <https://mybinder.org/>
- [10] Google Colab <https://colab.research.google.com>
- [11] J. Atenas, L.Havemann & E. Priego (2015), *Open Data as Open Educational Resources: Towards transversal skill and global citizenship*, Open Praxis, 7(4)