

## Involving other communities through challenges and cooperation

---

**Clara Nellist\***

*LAL, Univ. Paris-Sud, CNRS/IN2P3, Université Paris-Saclay, Orsay, France*

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

### On behalf of the ATLAS Collaboration

The ATLAS collaboration has recently setup three projects targeting citizen science or specific communities :

The goal of the HiggsML project was to bring particle physicists and data scientists together through a challenge: compete online to obtain the best Higgs to tau tau signal significance on a set of ATLAS fully simulated signal and background. The challenge ran from May to September 2014, drawing considerable attention. In total, there were 1785 teams that participated, making it the most popular challenge at the time on the Kaggle platform.

The ATLAS@home project allows volunteers to run simulations of collisions in the ATLAS detector. During the first year the community mostly consisted of software fans, who were attracted by the technical challenge and contributed greatly to the debugging through the message boards on the website. With the start of LHC, the number of people attracted for outreach reasons grew.

Higgs Hunters is the first Particle Physics project hosted on the web-based citizen science platform, Zooniverse. Volunteers are asked to scan ATLAS data and Monte Carlo events, looking for secondary vertices. Results will be compared to the ATLAS secondary vertex finding algorithm in the context of the search for long lived particles in supersymmetric models. To date, more than 8,000 users have taken part, classifying more than 900,000 interesting features in ATLAS event displays.

The setup, current success and future of each project will be reviewed.

PACS: 01. Communication, education, history, and philosophy; 01.50.Rt Physics tournaments and contests; 01.75.+m Science and society.

*The European Physical Society Conference on High Energy Physics  
22–29 July 2015  
Vienna, Austria*

---

\*Speaker.

## 1. Introduction

The ATLAS collaboration has recently organised and managed three projects targeting citizen science or specific communities. These projects are the Higgs Machine Learning challenge, ATLAS@Home and Higgs Hunters. Each of these projects allow the public to be more actively involved with the scientific process used within the ATLAS collaboration.

## 2. Higgs Machine Learning

The aim of the Higgs Machine Learning (HiggsML) [1] event was to benefit from the knowledge of the machine learning community and to engage the public with the type of analysis work required to discover the Higgs boson. The symbolism of collaboration between the two communities was embodied in the poster for the challenge, shown in fig. 1.

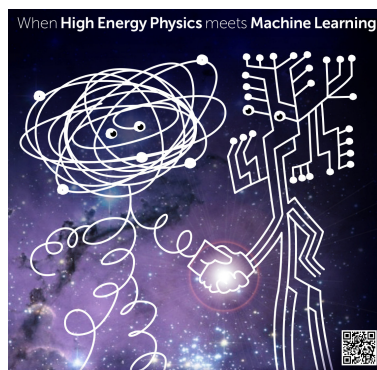


Figure 1: The HiggsML poster advertising the challenge. High energy physics is anthropomorphised as the character on the left, while the figure on the right represents the machine learning community.

The challenge ran from May to September 2014 on the Kaggle Platform [2] with a final total of 1,785 teams, 1,943 players and 35,772 entries. At the end of the challenge, it was Kaggle's most popular event so far. As is usual for challenges on Kaggle, teams with the highest score received monetary prizes with \$7000 for the first place, \$4000 for second and \$2000 for third. A *HEP meets ML* award was given to one team for providing XGBoost (eXtreme Gradient Boosted), a parallelised software to train boosted decision trees, which was used effectively by many of the other competitors.

The physics signal used was the Higgs boson decaying to two tau particles, which was simulated by the ATLAS experiment. Briefly, the goal within the challenge was to sort into two groups events that were due to a Higgs boson, known as the signal, and everything else, called the background. Illustrated as an example in fig. 2, there are a number of different ways the signal (in red) and background (in blue) could be sorted, as shown by the red lines. Either two straight cuts could be applied (left) a single straight line at an angle (centre) or more complicated cuts could be applied (right).

The major outcome from the challenge from the physics perspective was that the significance of the dataset increased by 20% with respect to the traditional HEP tool, Toolkit for Multivariate Analysis (TMVA) [3]. However, it was also a very fruitful collaboration established between

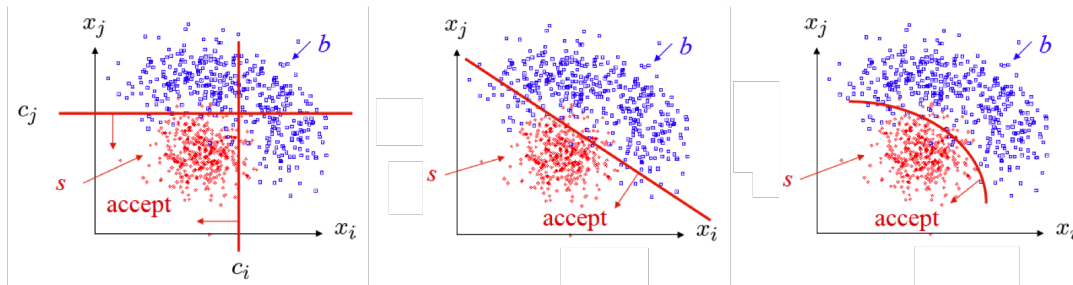


Figure 2: Three sketches illustrating how various cuts can be applied to separate signal (red) and background (blue).

ATLAS physicists and members of the machine learning community. New techniques and new software came from these collaborations, which have already been useful for ATLAS physicists. Just as importantly, these collaborations can be utilised for future developments that go beyond the scope of the challenge.

A workshop was held at CERN, Geneva in May 2015 to discuss the lessons learnt from the challenge and included presentations from the winner and the *HEP meets ML* award winners.

Although the HiggsML challenge on Kaggle has finished, the dataset used in the challenge, which contains approximately 800,000 events, will remain on the CERN Open Data Portal with a citable d.o.i.<sup>1</sup>. This will allow the dataset to be used for purposes outside of the challenge, such as teaching students how to analyse data from a high-energy physics experiment. Inspired by the HiggsML challenge, the LHCb collaboration have recently released their own challenge to ask people to identify a rare decay phenomenon [4], while ATLAS and also CMS are considering launching another challenge in the future.

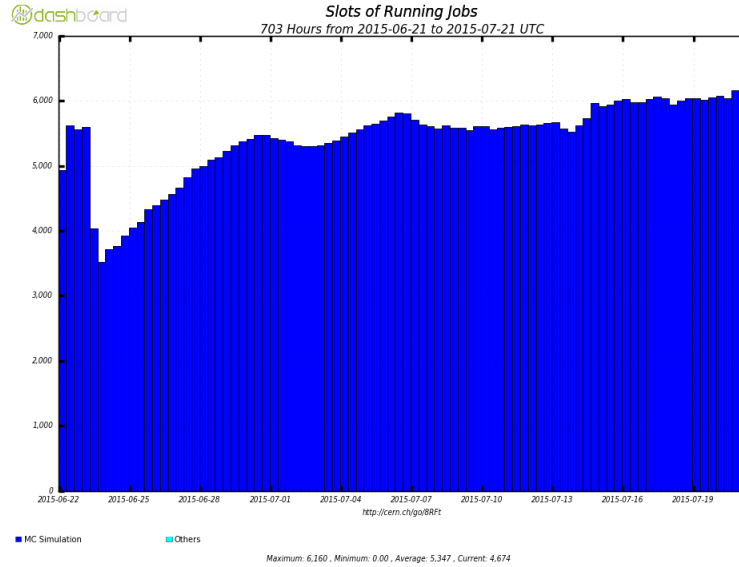
### 3. ATLAS@Home

The ATLAS@home project [5] allows volunteers to run simulations of collisions in the ATLAS detector using the idle processing time of their home computers. For the first year, the community consisted of enthusiastic software-savvy people who were attracted by the technical challenge and contributed to the debugging of the software via the message boards on the website. Indeed, many of these users found the project with no advertising at all. With the restart of LHC for Run 2 in 2015, the number of people attracted due to outreach reasons is growing. No knowledge of particle physics is required for a user to run the project, but for those interested in the physics processes simulated in ATLAS@Home a public page with further information is provided.

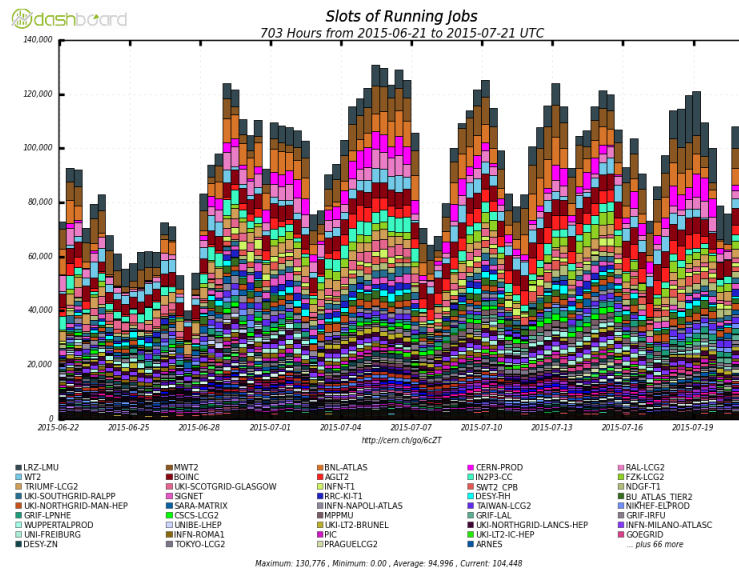
To participate in the project, volunteers download a free program called BOINC (Berkeley Open Infrastructure for Network Computing) [6], the platform that creates a virtual machine on the users computer for the ATLAS@Home project to run on. Each bundle of simulations is called a *workunit* on the BOINC site, or a *job* to the ATLAS physicists submitting them for processing. Depending on the processor speed of the users computer, each workunit can take approximately 1-2 hours to run. The number of jobs running on the BOINC platform as a function of time can

<sup>1</sup><http://opendata.cern.ch/collection/ATLAS-Higgs-Challenge-2014>

be seen in fig. 3a, while a comparison to other sites processing ATLAS jobs is shown in fig. 3b. The contributions of the BOINC platform of running ATLAS simulations is equivalent to a large University. A common complaint from volunteer users of ATLAS@HOME is that there are not enough workunits for them to run, which is indicative of their enthusiasm for the project.



(a)



(b)

Figure 3: Histograms showing (a) ATLAS jobs running on BOINC site and (b) all ATLAS Simulation Jobs.

To offer an additional incentive to users, various badges are available to earn from contributions of computer processing time, shown in fig. 4; this is a common feature on the BOINC platform. Rewards are for different contributions. The badges denoting quarks are available for

POS (EPS-HEP2015) 353

being in the top Nth percentage of the users contributing the most Recent Average Credit (RAC) where the first 1% of users are awarded a top quark. Leptons are earned for perseverance and after 180 days of contribution, users are given a tau neutrino award. The bosons are awarded for years of involvement, with the Higgs boson being awarded after 5 years.



Figure 4: Various badges for users of ATLAS@Home to earn through their contributions.

#### 4. Higgs Hunters

Higgs Hunters [8] is the first Particle Physics project hosted on the web-based citizen science platform called Zooniverse. This platform uses the efforts of volunteers to help researchers deal with the huge amount of data that they have produced to engage the public in the analysis and to understand which parts may be more suitable to direct human study, rather than a machine algorithm. So far, there have been greater than one million users worldwide, with 20 projects and 60 papers published. The Higgs Hunting project went live on the Zooniverse website in November 2014.

Volunteers are asked to look at events from measured ATLAS data and simulated events (called Monte Carlos) to search for secondary vertices, as illustrated in fig. 5. Also known as an Off-Centre Vertex, they are produced when a particle travels a measurable distance from the primary proton-proton collision point and then decays into subsequent particles. In this case the secondary vertex (or off-centre vertex as it is also known) could be from the decay of an exotic long-lived Higgs boson, a well-motivated model of physics beyond the Standard Model. To date, more than 8,000 users have taken part and have classified over 900,000 interesting features in 85,000 ATLAS event displays. Results from the project will be compared to the ATLAS secondary-vertex-finding algorithm in the context of the search for long lived particles in supersymmetric models to see if users found unusual or exotic events not found by the algorithms.

##### 4.1 Examples of interesting events found

Figure 6 shows two event displays which are examples of possible interesting events found by users of the Higgs Hunters website. The first, shown in fig. 6a is a possible beam halo event. The long red line, which doesn't cross the centre of the image, is suspected to have come from particles

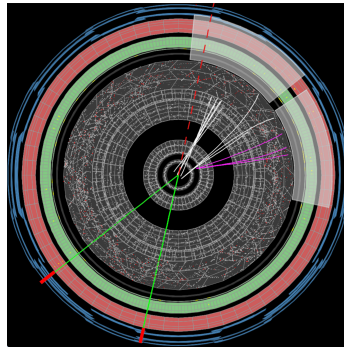
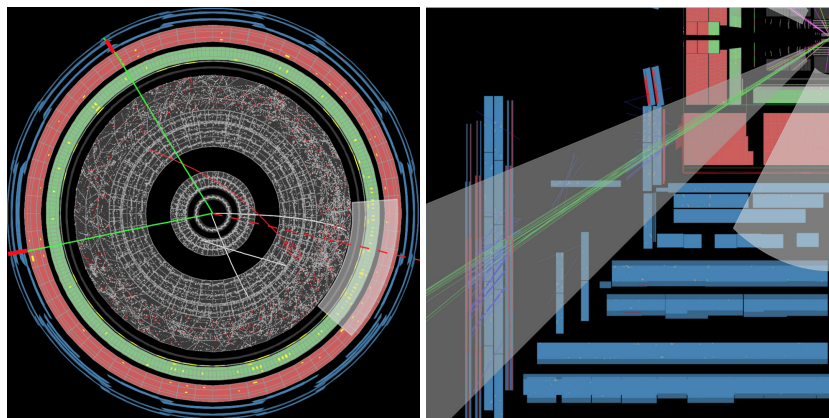


Figure 5: An example of an event display showing an off-centre vertex.

that were not created in the collision of two protons at the centre of the detector, but instead have been swept along with the LHC beam. The second event display, in fig. 6b was suspected to show a high-energy muon jet, but was later determined to be a high-energy particle that creates a jet of particles at the edge of the calorimeter and looks like a jet of muons. This is known as a ‘punch-through’ event.



(a) Possible ‘beam halo’.

(b) A ‘punch-through’.

Figure 6: Two examples of interesting event displays discovered by volunteers of Higgs Hunters.

## 5. Communicating the projects

To encourage greater participation in the three projects, communication to a wider audience is required. Social media platforms can be used to inform the public about these various projects and encourage participation and discussion.

An example is shown in fig. 7, which is a screenshot from a post during the CERN Ask Me Anything (AMA) event on Reddit in June 2015. In this post a member of the public learns that simulations for the LHC and ATLAS are run externally and asks how they themselves can help. The screenshots in fig. 8a and fig. 8b show examples of posts made on the official ATLAS Twitter account, which has a primary audience of over 30,000 followers, to advertise the HiggsML challenge and Higgs Hunters project respectively.

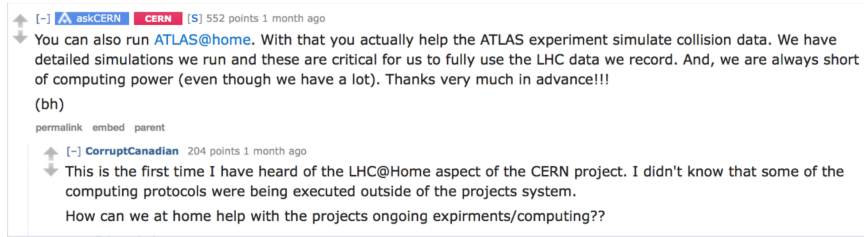


Figure 7: Screenshot from a post during the CERN Ask Me Anything (AMA) event on Reddit in June 2015.

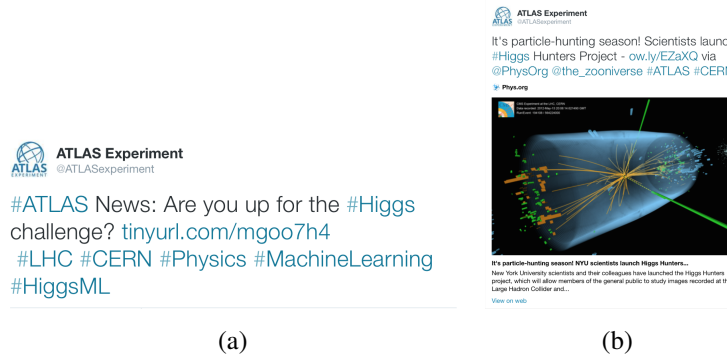


Figure 8: Screenshot from the ATLAS Twitter account advertising (a) the HiggsML Challenge and (b) the Higgs Hunters project.

## 6. Conclusion

Three projects which involve outside communities through challenges and cooperation have been presented. These projects, the Higgs Machine Learning challenge, ATLAS@Home and Higgs Hunters each have been shown to engage the public in the scientific process, benefitting the particle physics analyses within ATLAS whilst also providing an active forum for outreach.

## References

- [1] The HiggsML homepage: <http://higgsml.lal.in2p3.fr>.
- [2] The ‘about’ page on the Kaggle website: <https://www.kaggle.com/about>.
- [3] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, “TMVA: Toolkit for Multivariate Data Analysis,” PoS A CAT 040 (2007) [physics/0703039].
- [4] The Flavours of Physics homepage: <https://www.kaggle.com/c/flavours-of-physics>.
- [5] The ATLAS@Home website: <http://lhathome.web.cern.ch/projects/atlas>.
- [6] D. Anderson 2004 Proc. 5th IEEE/ACM Int. Workshop on Grid Computing, GRID 04. pp 4-10
- [7] XGBoost Documentation: <https://xgboost.readthedocs.org/en/latest/>
- [8] The Higgs Hunters website: <https://www.higgshunters.org>.