PROCEEDINGS OF SCIENCE

The REINFORCE citizen science project

C.Kourkoumelis*

Department of Physics, National and Kapodistrian University of Athens, University Campus,15784 Zografou, Athens ,Greece

E-mail: ckourk@cern.ch

The REINFORCE EU (Research Infrastructures FOR Citizens in Europe) was a three-year-long SwafS project which engaged citizens in active collaboration with the scientists working in large research infrastructures across Europe. The overall aim was to bridge the gap between them and reinforce society's science capital. The citizen scientists had at their disposal data from four different "discovery demonstrators" hosted on the online Zooniverse platform.

The demonstrators asked for the citizen contribution to front-end research such as: gravitational wave astronomy, deep sea neutrino telescopes, particle search at CERN and cosmic rays. The task of the citizens was to help the scientists to optimize the detectors and/or the reconstruction algorithms.

The focus of this contribution is on the demonstrator titled "Search for New Particles at CERN", where citizen-scientists visually inspected events collected by the ATLAS detector at LHC and searched for signatures of new particles.

The results of 360,000 classifications which show that citizen scientists can carry out complicated tasks responsibly and contributed in the ATLAS searches are presented as well.

This contribution is dedicated to the memory of late Prof. S. Katsanevas, the project's coordinator and former Director of EGO.

The European Physical Society Conference on High Energy Physics (EPS-HEP2023)

21-25 August 2023

Hamburg, Germany

* Speaker on behalf of the REINFORCE consortium

PoS(EPS-HEP2023)641

1. Introduction

The REINFORCE (Research Infrastructures FOR Citizens in Europe) was a three-year-long SwafS project [1] which aimed to engage citizens in the research performed at Large Research Infrastructures through citizen-science activities, in order to increase science-awareness and to positively influence attitudes towards it. The goal is to engage non-expert volunteers in citizen-science projects, ranging across all age groups and in high numbers - with no prior computing knowledge requirement- in active collaboration with the scientists across Europe. The project's design is built around the four so called "demonstrators", where research data are offered in order to attract and motivate volunteer citizens to become a part of, and contribute to, frontier research science topics. The demonstrators are hosted in the Zooniverse platform [2], the most widely used citizen science platform with over 2,5 M users. The four demonstrators and the accompanying data are the following:

- GWitch Hunters [3]: Data from VIRGO, a giant laser interferometer designed to detect gravitational waves (GW), located close to Pisa in Italy. The aim is to get citizens help the scientists in the search for GW by recognizing and classifying different noise patterns.
- Deep Sea Explorers [4]: Data from the KM3NeT, neutrino telescopes installed in the depths of the Mediterranean sea, aiming to detect ultrahigh energy neutrinos from the cosmos. Ctizens should hunt for "what is not a neutrino", namely should study sources of bioluminence and acoustic noise detected by hydrophones in deep sea.
- New Particle Search at CERN [5]: Data from the the world's most powerful collider (the Large Hadron Collider; or LHC) for frontier research in high-energy physics, More details are given in the next sections.
- Cosmic Muon Images [6]: Data from atmospheric muons telescopes for geoscience and archaeological studies.

In addition to the above four demonstrators, an important objective of REINFORCE managed to extend the data handling and research to the scientific community which is sense-disabled (especially visually-disabled) and senior citizens. It also emphasized the interplay of art and science, organizing exhibitions and contests.

2. The "New Particle at CERN" demonstrator

2.1 Description of the demonstrator

The "New Particle Search at CERN" is a demonstrator which engages citizens in the state-of-the-art particle research performed at the LHC of CERN in the quest for the understanding of the ultimate structure of matter. The demonstrator is based on data collected by the ATLAS experiment [7] which are produced by high-energy proton-proton collisions at the LHC. The volunteer citizens are motivated to perform a visual inspection of data samples consisting of "events", namely the registered products of proton-proton collisions. This way, they contribute to the search for yet undiscovered hypothetical particles predicted by theories Beyond the Standard Model (BSM) [8].

To enable the searches, the demonstrator adopted a three-stage architecture. The first two stages are based on selected samples of simulated data in order to train citizens, but also to allow for a quantitative assessment of their performance and a comparison with specially developed automated algorithms. The third stage of the demonstrator is a “discovery” one, employing datasets of real events from the ATLAS Open Data Set Release [9].The third stage provides two research paths: (a) study of Higgs boson decays to two photons, one of which could be converted to an electron-positron pair by interaction with detector material, and (b) search for yet undiscovered neutral long-lived particles, predicted by certain theories of the BSM. The citizen research involves the identification of specific “signatures” which are produced from the decay of these new long-lived particles. These decay products could originate from displaced vertices (DVs), namely vertices -formed by two or more tracks - that are displaced with respect to the main collision point of the two protons. The demonstrator has developed visual analysis tools that allow the citizens not only to classify static images –in order to recognize the DVs- but also to interact with the event displays, select specific tracks and calculate kinematical quantities characteristic of the sought-after particles.

In Stage 1, which is hosted entirely on Zooniverse, citizens are trained to recognize DVs in a high purity sample of simulated data, corresponding to the various scenarios of new particles with displaced vertices. The users only inspect stationary images of the traces which charged particles leave in the inner part of the ATLAS detector. The users look for tracks which intersect at a point other than the main interaction point. They need to inspect both views of the inner detector to be able to properly recognize track intersection. In the two different views/projections of the inner detector which are depicted in Figure 1, the tracks are coloured so that a user may identify the same track in both views. The user is then asked to spot the displaced vertices in both views using the mouse, and the answer is internally assessed by Zooniverse based on the truth information which is also provided to the platform.

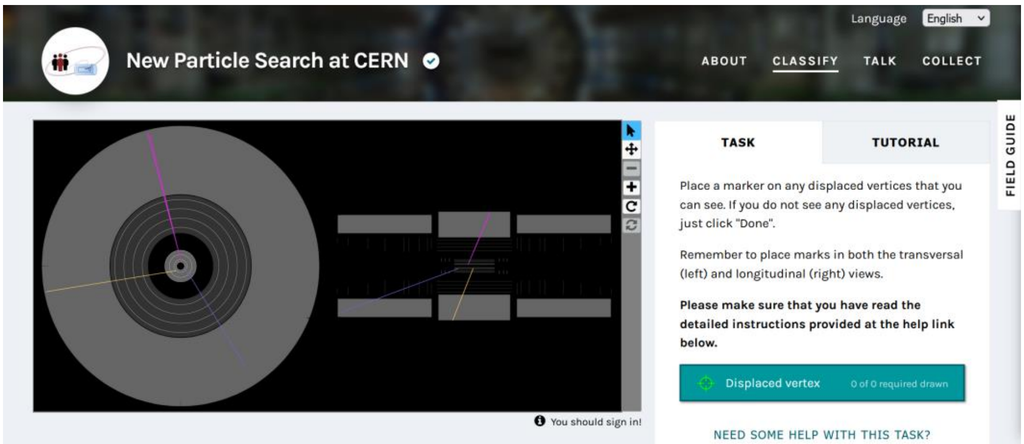

Figure 1: The First Stage of the demonstrator on Zooniverse

For Stages 2 and 3 the users are directed from Zooniverse to the HYPATIA [10] event display tool which shows two views of all ATLAS detectors, together with a table of relative information on the collision products. The users can interact with the tool - instead of simply examining static images and locating DVs - in order to perform in-depth analysis of the events. In the context of the demonstrator, significant additions were made to HYPATIA, in order to incorporate the functionality necessary to enable the citizens to perform their analysis in each stage. In addition, a new event format was developed which includes additional information such as the display of DVs.

In Stage 2 the users are asked to identify certain particle types which are useful for the next “discovery” stage. They look for signatures of muons, electrons, photons and converted photons. The HYPATIA particle information table shown in Figure 2, displays momentum, charge and direction information for each track or cluster of the event. The user can identify tracks/clusters, using the button that corresponds to the type of track or cluster that they have determined it belongs to (electron, muon, photon or converted photon). When the user clicks the ‘Next’ button to display another event, the selections made are stored in a private database for later processing. Since the sample consists of simulated events, the particle generation identification is known and can be compared to the user classification, in order to determine the validity of the choice made.

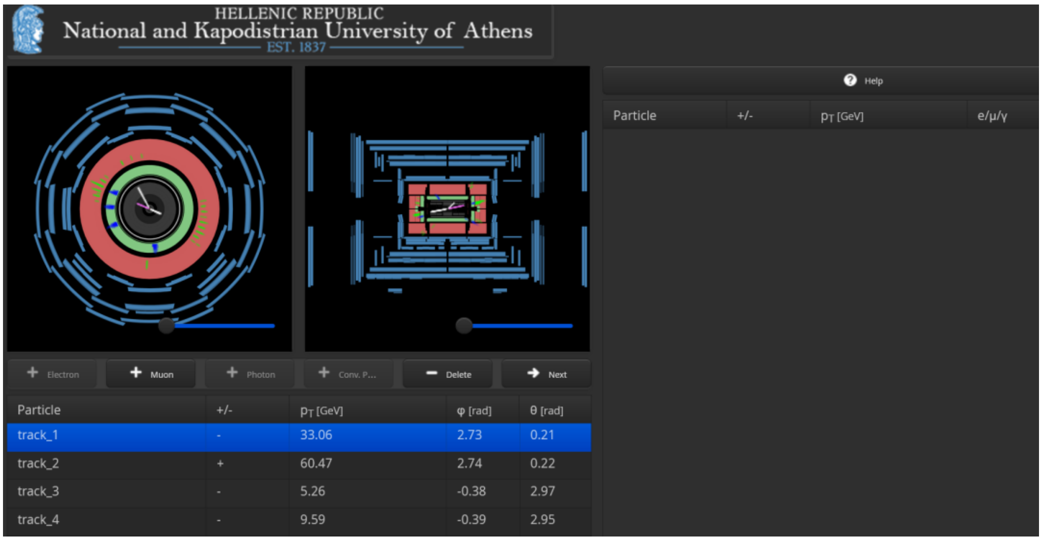

Figure 2: The second Stage of the demonstrator on HYPATIA.

In the ‘Higgs boson’ part of Stage 3, the user can indicate that a cluster belongs to either a photon or a converted photon that could originate from Higgs boson decays. HYPATIA automatically calculates the invariant mass when the user selects a pair of photons. In addition, the user can rate the event (from one to five stars) based on the instructions given on the platform. In this study, users selected 94% of all photon-pair masses in the 106-160 GeV mass range. Distinguishing the Higgs boson signal, which lies on top of the background, requires a much larger dataset than the one available to the users, and a sophisticated statistical analysis of the data.

In the 'Long-Lived-Particle (LLP) Hunting' part of Stage 3, users searched for long-lived particles predicted by some BSM models, through the identification of secondary vertices of particle decays. Following the vertex identification, they were asked to look for muons originating from the displaced vertex and to examine certain kinematic variables, which could be characteristic of a candidate long-lived particle. Users were also asked to rate each event from one to five stars, according to how similar they believed it to be, to one of the sought-after decays.

2.2 Comparative study of user classification results

In Stage 1 the quality of the citizens' work is assessed and compared against a specially developed automated algorithm which detects the presence of displaced vertices in the events. The algorithm extrapolates the tracks and looks for their intersection, using only detector information that is also available to the users. The overall goal was to compare the DVs identification efficiency of users to that achieved by the algorithm. The classification efficiency, is defined as the fraction of correctly identified DVs over the total number of DVs in the events, which the users have processed. The results from 180,000 classifications provided by the users were analysed in multiple ways. The most interesting results were obtained by considering the 'user consensus' as an identification criterion. The "consensus" is based on events that have been classified by at least 10 different users and at their combined result. The identification efficiency of the user consensus is found to be about 93% on average, depending on the view of the ATLAS detector provided (89% in the transverse detector view, 96% on the longitudinal detector view).. This efficiency is very close to the respective identification efficiency obtained by the dedicated automated algorithm, which was 94%.

In Stage 2 since the sample consists of simulated events, the particle generation identification is known and can be compared to the user classification, in order to determine the validity of the choice made. The users made a total of 37,000 classifications 80.4% being correct. Each identification answer was also compared to the output of a Machine Learning (ML) algorithm. Table 1 summarizes the results in terms of the efficiency and purity for each type of particle for both citizens and the ML algorithm. The comaprison shows that the users answers are not very far from the dedicated ML ones.

Particle type	User efficiency (%)	ML efficiency (%)	User purity (%)	ML purity (%)
Electron	90	97	88	78
Photon	96	97	80	78
Conv. photon	62	95	35	81

Table 1: Comparison of the user particle identification efficiency/purity with the ML algorithm results

2.3 User results from the discovery phases

In the two discovery parts of stage 3, where real data are used, the citizen scientists showed enthusiasm to search for event types that could correspond to "new particle" candidates. They were able to recognize "interesting" events and rate them accordingly.

PoS(EPS-HEP2023)641

In the 'Higgs boson' study the users rated 1,156 events with five stars, with most events receiving low star-ratings as they did not contain unusual Higgs candidate decays. Of those 1,156 five-star rated events, only a few contain extra leptons, but the invariant mass of the photons (or converted photons) is outside the mass range of the Higgs boson mass, therefore those events cannot be attributed to complex Higgs boson production and decay mechanisms.

In the 'LLP Hunt' part a total of 81,894 classifications were performed on a sample consisting of 2,440 events. Events which satisfy all given requirements have a much higher probability to originate from a new long-lived particle. Since these particles are, as yet, undiscovered, the relevant candidates are extremely rare. The users only viewed 27 of them (users viewed each one multiple times) and correctly marked with five stars 23 of those events. Two of the five-star events were good candidates for new BSM long-lived particles (hypothetical supersymmetric particles). The events were scrutinized by our team and further information (which was not available to the volunteers) was inspected. It turns out that, most probably, both events are due to the expected background. Moreover, a number of volunteers did isolate 24 events with more than one muon in the DV and posted their findings in the Talk forum, where they were discussed by the wider community. After further investigation by our team, it was revealed that these events were either due to the interactions of a known particle with the detector or faults in the reconstruction of particle traces.

2.4 Interaction with the users

From the launch of the "New Particle Search at CERN" on the 19th of October, 2021, until the 30th of November, 2022 the responsible team received and replied about 3,000 Talk messages (seven messages per day on average) posted by users on the talk forums of the Zooniverse platform. Volunteers have been communicating to us technical questions regarding the stages of the project (mostly during the first months), questions on physics related (or not) to the tasks provided to them, their results and observations, as well as interesting suggestions and features they would like us to add.

Based on user results, along with the enthusiastic and useful comments and suggestions received by a large number of citizens in the Zooniverse Talk section, the New Particle at CERN demonstrator platform was quite successful. It showed that citizen scientists can carry out complicated tasks responsibly, with a performance comparable to that of a purpose-built machine-based algorithm, and can identify interesting patterns or errors in the reconstruction, in individual events. Finally, the demonstrator showed that the statistical combination of user responses (user consensus) appears to be quite a powerful tool that can be further considered and exploited in fundamental scientific research.

3. Conclusion

About 12,000 citizens are involved, with 1,600,000 classifications in all four demonstrators. All four demonstrators are available in English, Spanish and Greek. In addition the GWitch Hunters is also available in Italian. Comparative studies show that the power of citizen's performance (often similar to ML algorithms).

Part of the data of all demonstrators gave been sonified by the Argentinian CONICET partner using their SonoUno [11] software. The sonification hands-on results show that visually impaired citizens can equally participate in the analysis effort. Moreover, the project's site houses a platform of all artistic interventions of the project.

As part of REINFORCE's sustainability we will continue gathering data and monitoring user responses. It is possible that this could provide us with some more insights on user abilities and strengths. Beyond REINFORCE, a future project specifically targeted to those areas could provide more information and suggest new ways of utilizing citizen science in the field of particle physics.

4. Acknowledgements

This project has received funding from the European Union's Horizon 2020 project call H2020-SwafS-2018-2020 funded project Grant Agreement no. 872859

References

[1] https://reinforceeu.eu/

[2] https://www.zooniverse.org

[3] https://reinforceeu.eu/demonstrators/ gwitchhunters

[4] https://reinforceeu.eu/demonstrators/ deep-sea-explorers

[5] https://reinforceeu.eu/index.php/demonstrators/ new-particle-search-cern

[6] https://reinforceeu.eu/demonstrators/ cosmic-muon-images

[7] https://atlas.cern/

[8] ATLAS Collaboration, *"Search for massive, long-lived particles using multitrack displaced vertices or displaced lepton pairs in pp collisions at* $\sqrt{s}$ *= 8 TeV with the ATLAS detector*", Phys. Rev. D92, 072004 (2015)[https://arxiv.org/abs/1504.05162]

[9] ATLAS Collaboration, *Review of the 13 TeV ATLAS Open Data release*, PUB-OTRC-2020-01, 22nd January 2020

[10] C Kourkoumelis and S Vourakis, *HYPATIA-an online tool for ATLAS event visualization,* Phys. Educ. 49 (2014) 21 doi:10.1088/0031-9120/49/1/21 http://iopscience.iop.org/0031-9120/49/1/21/

[11] https://www.reinforceeu.eu/about/sonification-increasing-senses-increasing-inclusion