

The ESCAPE Data Science Summer School 2021

Thomas Vuillaume¹, Mathilde Hubert¹, Kay Graf^{7,8}, Eduardo Rodrigues², Vincent Poireau¹, Arturo Sánchez Pineda¹, Alberto Iess³, Maximilian Nöthe⁴, Martino Sorbaro⁵, Tamas Gal^{6,7}, Rachael Ainsworth⁸, Karl Kosack⁹, Enrique García¹ and Axel Donath¹⁰

¹*Univ. Savoie Mont Blanc, CNRS, Laboratoire d'Annecy de Physique des Particules - IN2P3, 74000 Annecy, France; vuillaume@lapp.in2p3.fr*

²*Oliver Lodge Laboratory, University of Liverpool, Liverpool, United Kingdom*

³*Scuola Normale Superiore, Pisa, Italy*

⁴*Astroparticle Physics, TU Dortmund University, Germany*

⁵*AI Center, ETH Zurich, and Institute of Neuroinformatics, University of Zurich and ETH*

⁶*Erlangen Centre for Astroparticle Physics (ECAP)*

⁷*Friedrich-Alexander-University Erlangen*

⁸*University of Manchester, Manchester, United Kingdom;*

⁹*IRFU, CEA Paris-Saclay, Université Paris-Saclay, France*

¹⁰*Center for Astrophysics | Harvard & Smithsonian, Cambridge MA, US*

Abstract. The goal of reaching an Open Science ecosystem in the European Open-Science Cloud (EOSC) is not possible without training of early career scientists. In particular, the creation and maintenance of high-quality and open software need special consideration. This is tackled within the ESCAPE H2020 project through thematic training events: the ESCAPE Data Science Summer Schools. During these schools, young scientists in the field of astronomy, astro-particle and particle physics are taught the necessary ingredients for their software to become a part of open science by experienced code custodians. The 2021 school's edition was a continuation of the ASTER-ICS/OBELICS summer schools previously organised at LAPP (Annecy, France) but has been re-organised as an online event due to the COVID crisis. It witnessed more than 1000 registered participants, showing both the need and importance of such training events. Following the FAIR paradigm and as an example of good practices in code development, the full information of the school are openly available online, including scientific programme, agenda and links to all contributions (software repository, notebooks, contributions, presentations and recordings). In this contribution, we review the goals, organisation and participants feedback of such an event to present a return of experience for future similar events.

1. The ESCAPE H2020 project

ESCAPE¹ addresses the Open Science challenges shared by ESFRI facilities (CTA, ELT, EST, FAIR, HL-LHC, KM3NeT and SKA) as well as other Pan-European research infrastructures (CERN, ESO, JIVE and EGO) in astronomy and particle physics. ESCAPE actions focus on developing solutions for the FAIRness of large data sets handled by the ESFRI facilities. A key aspect in order to secure the accomplishment of the ESCAPE objectives is the training of its communities, in particular early-career scientists by experienced code custodians.

2. The ESCAPE data science summer school: programme and objectives

The ESCAPE data science summer school is a continuation of the ASTERICS summer schools (2017, 2018, 2019) and has been held online from 7 to 18 June 2021. The aim of the school is to provide theoretical and hands-on training on Data Science and Python development for astrophysics, astroparticle physics and particle physics. It has been designed around the partners' needs and their open science developments. The lectures are organised around 6 blocks: Coding Environment, Version control, collaborative and continuous development and integration, Profiling, debugging and optimising, Scientific Python libraries, Introduction to Machine Learning and an introduction to other tools such as Julia and Spark. Only a basic knowledge of the Python programming language was requested. The lectures are then designed and given by experienced and talented developers from partners. The school targets any software engineer, data scientist or interested students from the astronomy, astroparticle and particle physics communities. The stakeholders are composed of the Pan-European Research Organisations, the ESFRI projects, the ESCAPE partners² and small and medium companies and the Scientists Communities.

3. Organisation of the school

3.1. Free and open teaching for open science

Being online allowed us to propose a completely open and free-of-charge event. Registrations were done through the Indico platform with a simple registration form, without any selection. Following the FAIR paradigm and as an example of good practices in code development, all the lectures material (slides, code, notebooks, recorded lectures) was and still is freely accessible to anyone through open licenses, opening the possibility to modify it and re-use it in other contexts. It has been archived on Zenodo for long-term preservation (Vuillaume et al. 2021).

3.2. Online tools and resources

In order to recreate a participative environment online, we integrated several solutions and online tools to facilitate the communication between participants and tutors. Here

¹<https://projectescape.eu/>

²<https://www.projectESCAPE.eu/partners>

is a list of the tools and platforms used during the school to organize and share the material, as well as communicate with all school participants.

Indico³: provides detailed program and the main information about the school organization. It has been used to manage participants' inscriptions.

GitHub⁴: Lecturers contributed their material directly to this repository. The issues have been used for participants to report technical issues they could experience.

Lectures portal⁵: a web portal providing a friendly experience to participants to find all the school material organized per lectures.

YouTube channel⁶: The lectures were broadcasted live on YouTube where anyone could attend. Not only YouTube is a well known and easy to use platform for anybody, it eased the recording of the lectures that were readily available to be re-watched offline. Moreover, YouTube allows viewers to stop the video, even during a live session, thus allowing them to make sure they understood correctly a specific point or some manipulation tutors did and quickly go back to the live, something not possible with other visio-conference tools.

Slack⁷: the main communication tool where all announcements were made. It was used during the lectures by participants to ask questions that were either answered directly, or conveyed to the lecturer for live feedback. It was accessible only to registered participants to the school.

GatherTown⁸ is a platform that provides a virtual world to ease visio-communication. It was used mainly during coffee breaks or small group sessions.

Twitter was used to promote the school, the courses and the teachers before and during the event with the specific hashtag #ESCAPESchool21.

LinkedIn was used to promote the school, the courses and the teachers before and during the event. <https://www.linkedin.com/company/projectescape>

4. Results

Two surveys using the open tool LimeSurvey hosted at LAPP were conducted among registered participants before and after the school, in order to better understand their background, knowledge and origins, to better adapt and modulate the content and then to get their feedback. More than 400 participants participated in both of these surveys.

As a result of the open and free registrations, the event was truly international with all continents represented. Also, the vast majority of participants (96%) was not directly linked to ESCAPE partner institutions, a percentage not achieved with in-person schools, presumably due to the associated travel costs. Participants are mostly students or young researchers as shown in figure 2. Although the gender gap still exists in the physics communities but we have a participation of more than 41% of women. The

³<https://indico.in2p3.fr/event/20306/>

⁴<https://github.com/escape2020/school2021>

⁵<https://escape2020.github.io/school2021/>

⁶https://youtube.com/playlist?list=PL513baTu76qPbYZ1RK0Yhr4P_wsI-84Hn

⁷<https://escape-data-school.slack.com/>

⁸<https://gather.town/app/Rww2ZWwsxiA2Usz3/ESCAPE%20school>

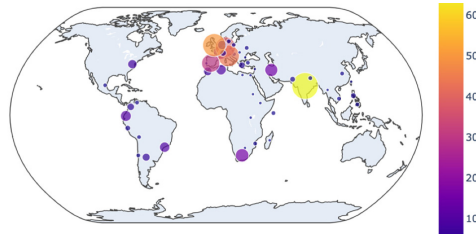


Figure 1.: Number of participants by country that replied to the school survey.

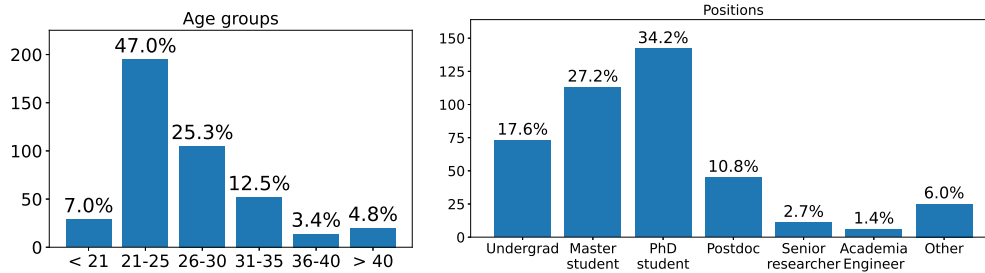


Figure 2.: Participants ages and positions distributions from the survey

participants were generally very satisfied with the school that globally met their expectations (fig. 3).

Acknowledgments. We gratefully acknowledge financial support from the European Union’s Horizon 2020 research and innovation programme under Grant Agreement no 824064, the local organizing committee and teachers who volunteered to share their knowledge.

References

Vuillaume, T., Nöthe, M., Peloton, J., Donath, A., Pineda, A. S., Rodrigues, E., Garcia, E., Kosack, K., Gal, T., Muite, B., Iess, A., Sorbaro, M., Beleites, C., Schnabel, J., & Ainsworth, R. 2021, Escape data science summer school 2021. URL <https://doi.org/10.5281/zenodo.5093909>

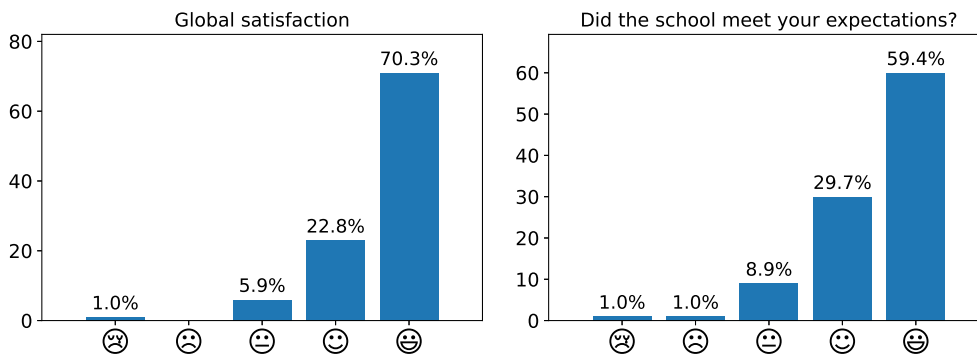


Figure 3.: A high level of satisfaction from the participants survey.