PROCEEDINGS OF SCIENCE

The Pierre Auger Observatory Open Data

V. Scherini[a,*] **and The Pierre Auger Collaboration**[b]

[a] *Università del Salento and INFN Lecce, Italy*

[b] *The Pierre Auger Observatory, Av. San Martín Norte 306, 5613 Malargüe, Mendoza, Argentina; http://www.auger.org*

Full author list: *https://www.auger.org/archive/authors_2024_11.html*

E-mail: spokespersons@auger.org

During almost 20 years of regular data acquisition, the Pierre Auger Observatory, the world's largest facility for measuring ultra-high energy cosmic rays, has collected a vast and diverse amount of data covering complementary research fields from astroparticle and fundamental physics to space weather science.

The Pierre Auger Collaboration has embraced the concept of open access to research data since its foundation. Since then, a gradual release process has been initiated, and a dedicated task force has been established to implement and sustain this effort over the long term. In line with the FAIR (Findable, Accessible, Interoperable, and Reusable) principles, the mandate of the task force includes the selection of data samples and the documented translation of in-house analyses codes into popular and available software for easy exploration and manipulation of the data.

The Pierre Auger Open Data Portal was launched in February 2021 and contains 10% of the cosmic-ray data and 100% of the atmospheric and space-weather data. Since its initial release, the Portal has been extended and updated with new data samples. It also includes a detailed catalog of the showers created by the highest-energy particles and an outreach section aimed at engaging the general public in cosmic-ray science. The foreseen increase of the fraction of released cosmic-ray data to 30% and the inclusion of new detectors will further boost the scientific community's interest in the Observatory's data and their use for education and outreach initiatives.

7th International Symposium on Ultra High Energy Cosmic Rays (UHECR2024)
17-21 November 2024
Malargüe, Mendoza, Argentina

*Speaker

ISSN 1824-8039. Published by SISSA Medialab.

PoS(UHECR2024)114

Figure 1: The Southern sky at the Pierre Auger Observatory. Photo credits to Steven Saffi.

1. Motivations and challenges

Data from the Pierre Auger Observatory [1], Fig 1, come from a variety of instruments and take many forms, starting from either raw experimental or simulated data through reconstructed data and higher level data generated by analysis workflows, all the way up to data presented in scientific publications. The data result from an enormous and long-term human and financial investments by the international community.

The Collaboration is committed to the public release and provides accompanying software tools to offer to a broader community, including professional and citizen scientists, a unique opportunity to explore and analyse the data at various levels of complexity. This is inspired by the FAIR (Findable, Accessible, Interoperable, and Reusable) principles [2].

The Collaboration approach to data open access [3] needs a complex and continuing effort based on offering:

- support and facilitation: a detailed explanation of detection techniques, data reconstruction and selection
- portable and flexible file format: the use of JSON (JavaScript Object Notation) and CSV (comma-separated values)
- analysis code and tutorials: Jupyter Notebooks in Python for easy manipulation of data.

A task for Open Data has been created under the responsibility of the Project Management to facilitate the effort, working in synergy with the related physics tasks. An efficient procedure has been designed to translate the simplified and portable format for Open Data from proprietary binary files. Data are processed with the most up-to-date reconstruction software, and changes are propagated into updates of the released data. The validation of the data samples and the codes is performed under the supervision of analysis and detector experts. The detailed accompanying documentation is published following the standard internal review.

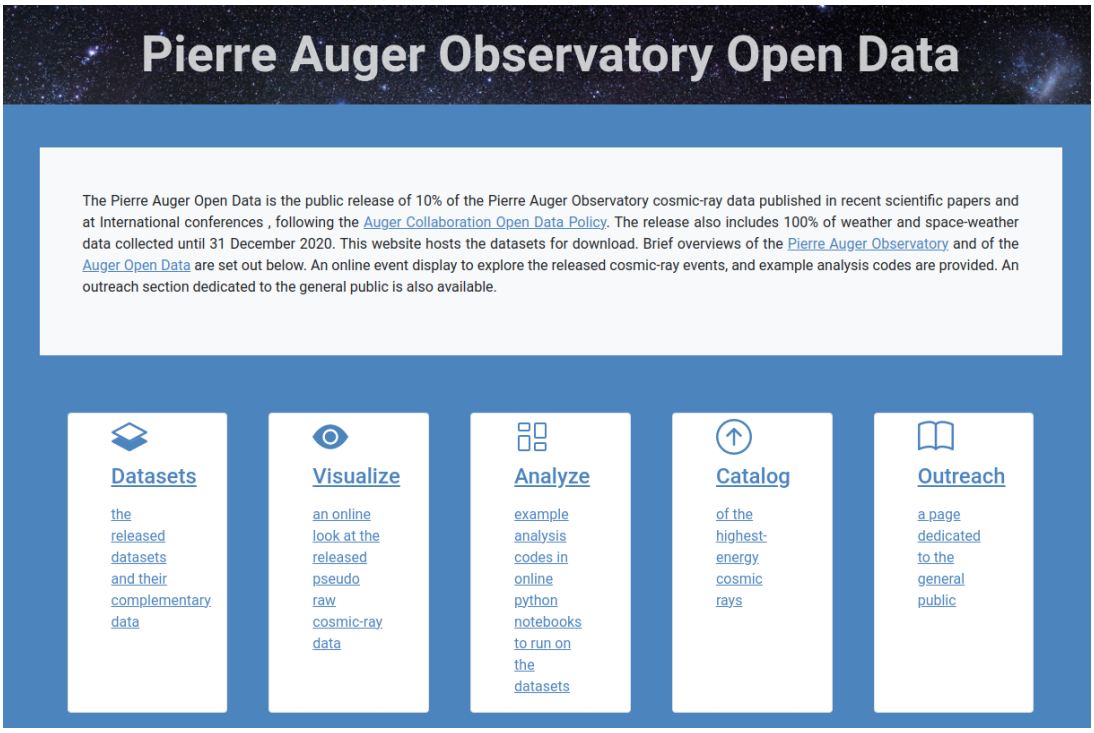

Figure 2: Screenshot of the main page of the Open Data Portal [4], with the links to the five different sections of the page. For more details see [8, 9].

2. The Open Data Portal

Following the Auger Collaboration Open Data Policy, the Open Data Portal [4] contains the public release of 10% of the Pierre Auger Observatory cosmic-ray data published in scientific papers and at international conferences and 100% of the atmospheric and space-weather data. Data are released under the CC BY-SA 4.0 International License at the Zenodo DOI [5]: https://doi.org/10.5281/zenodo.4487612. Brief overviews of the Pierre Auger Observatory and an online event display to explore the released cosmic-ray events and example analysis codes are provided. An outreach section dedicated to the general public is also available. The sections are visible in the screenshot of the portal displayed in Fig. 2. The portal evolution has been described in [6–8] and is dynamically tracked in [9].

Datasets

The Pierre Auger Open Data consist of a cosmic-ray dataset of more than 80 000 showers, measured with the surface detector (SD), and of more than 3 000 hybrid events, i.e. recorded simultaneously with the surface detector and the fluorescence detector (FD). The Open Data also include the counting rates of the surface detector stations, recorded with scalers and averaged over every 15 minutes from 2005 to 2020, and atmospheric data acquired with weather stations.

In addition, auxiliary data files are provided, namely the list of the positions of the surface detector stations and the pixels of the fluorescence detector telescopes, as well as the exposure of the surface detector arrays and the parameters needed to calculate the fluorescence detector acceptance.

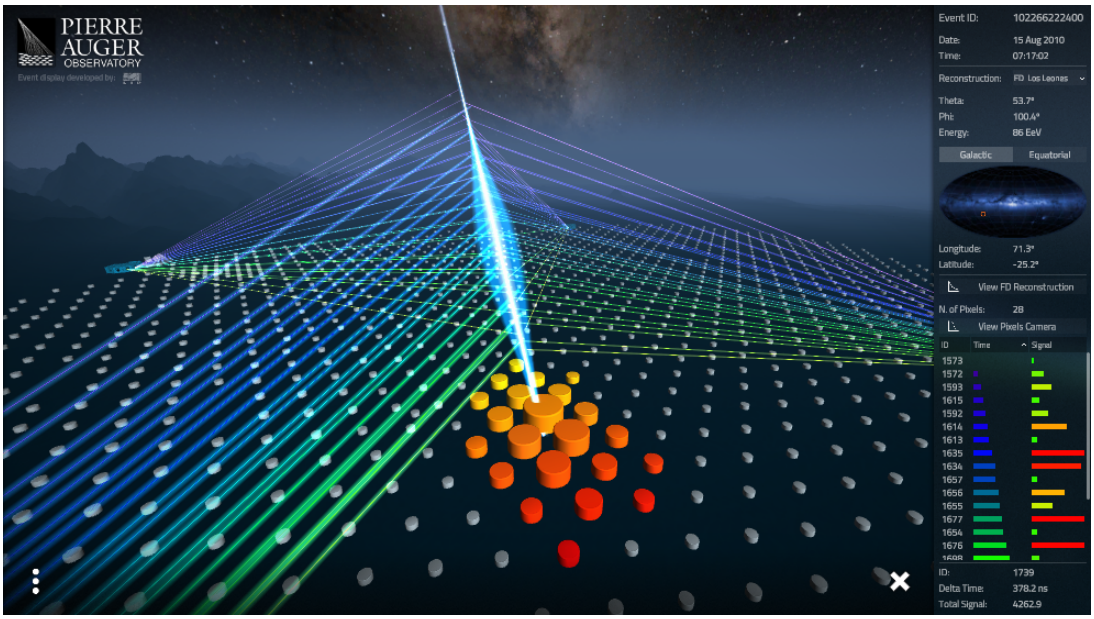

Figure 3: The highest-energy multi-eye hybrid event in the UHECR catalog [10], PAO100815 (id 102266222400): the reconstructed zenith angle is 54°, the energy 82 EeV. It triggered 22 stations of the surface detector and four fluorescence detectors.

The datasets are detailed as follows:

- 10% cosmic ray dataset:
  - ~ 80 000 selected events collected by the surface detector in the period 2004 - 2018 used in the ICRC 2019 in Madison, USA. The samples contain ~ 25 000 SD 1500 m array events above full efficiency threshold, with zenith angle < 80°, meaning 80% coverage of the sky, and ~ 55 000 SD 750 m events above 0.1 EeV and zenith angle < 40°
  - ~ 3 300 hybrid events, collected simultaneously by the surface and the fluorescence detectors and selected according to specific analyses.

  The selected events are available for download in JSON files containing calibrated data for each surface detector station participating in the event with their recorded photo-multiplier traces. Summary files in CSV format containing high-level info with reconstructed parameters are also provided

- 100% atmospheric data: local condition parameters such as pressure, temperature, humidity, and wind speed measured at the Auger site by weather stations and monitoring devices

- 100% scaler mode data: rates acquired by surface detector stations in low threshold particle counter mode for space-weather studies.

Visualization

A user-friendly interface is available for selecting and browsing each public event by specifying an event ID or a range of reconstructed variables, such as the energy or the zenith angle. The browser contains an immersive 3D animation from the arrival direction of the cosmic rays to the

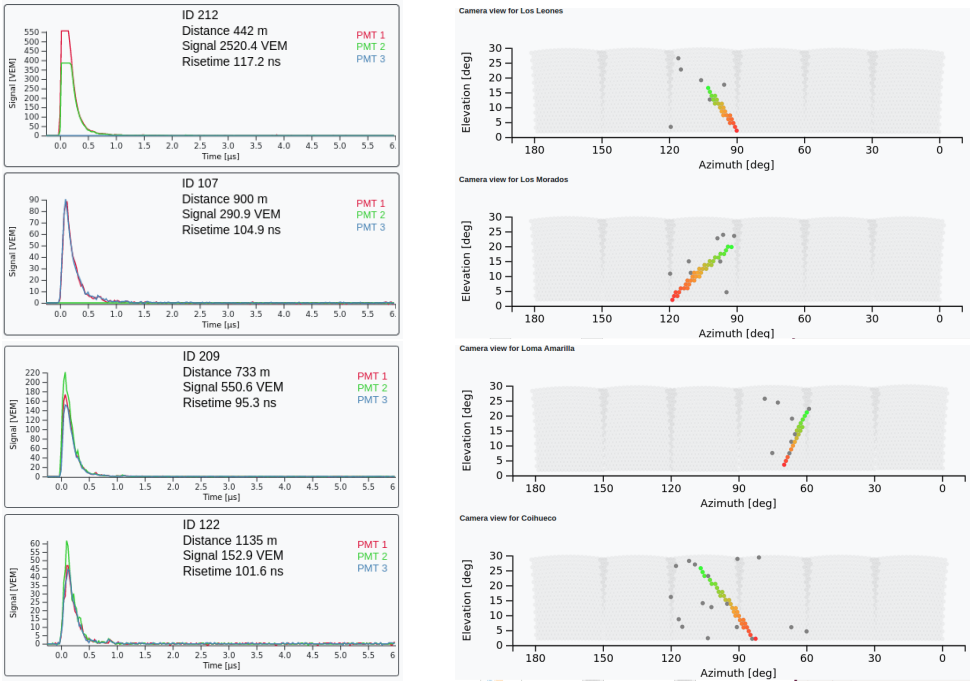

Figure 4: The highest-energy multi-eye hybrid event in the UHECR catalog [10], PAO100815 (id 102266222400): example of traces in the surface detector stations (left panels) and camera view of the fluorescence detectors (right panels).

detection of the created extensive air shower with the instruments of the Observatory, see Fig. 3. The traces recorded in the photo-multiplier tubes of two surface detector stations and the camera view of the triggered telescopes in the fluorescence detectors are displayed in Fig. 4, left and right panels respectively.

Analysis

The Open Data can be analyzed using the provided Python Jupyter Notebooks. Tutorial examples are provided in the Portal introducing the Python programming language and its use with the Open Data. More advanced analysis codes are simplified re-implementation of parts of analyses published by the Collaboration. All the Notebooks can be run online on a web browser via the Kaggle platform or downloaded together with the datasets. The graphical output of exemplary notebooks as energy calibration, spectrum, mass composition, and arrival directions is shown in Fig. 5.

Outreach

The Outreach section, aimed at a wider audience and translated into several languages, provides a unique opportunity to share the excitement of cosmic-ray physics with students, teachers and citizen scientists in the general public. Built in the same spirit as the research part, with the same data but in a simplified format, provides exemplary tutorial and analysis tools to understand and manipulate the released data. It invites people to explore data and use them for their own inquiry by developing original education and outreach activities.

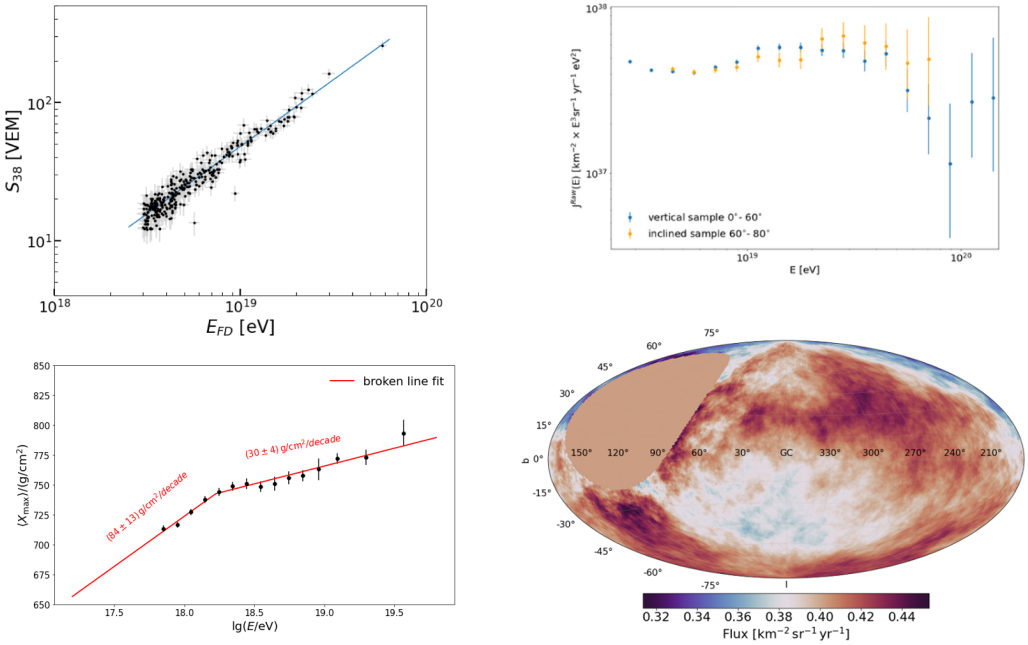

Figure 5: Graphical output of the analysis notebooks: energy calibration and spectrum (top), mass composition and arrival directions (bottom).

3. The UHECR Catalog

The events published in the catalog of the 100 highest-energy cosmic-ray events [10], with reconstructed energy between 76 EeV and 166 EeV, collected during Phase I of the Observatory's data taking (between 2004 and 2021) used in the study of the arrival directions of events above 32 EeV [11], along with the nine highest-energy hybrid events used for their calibration, are available for inspection and download in the UHECR catalog section of the Portal.

For these events, all details of the reconstruction are available in the event summary card, such as the Coordinated Universal Time (UTC), the energy, the zenith and azimuth angles, the declination and right ascension, the multiplicity of triggered stations, and, for hybrid events, the quantities measured with the fluorescence detector, such as energy and depth of shower maximum.

Additional features can be viewed: not only the footprint at the ground can be displayed, but also that on the shower plane, and besides the lateral distribution of the shower particles, the user can also see the time delays of the signals with respect to a plane shower front. The associated JSON files contain the calibrated traces for each photomultiplier tube in the triggered stations. Displays for the highest-energy hybrid event in the UHECR catalog are shown in Fig. 6.

The catalog demonstrates the quality of the data that lie behind measurements of the energy spectrum, the distribution of arrival directions, and the mass of the highest-energy cosmic rays that have been reported by the Pierre Auger Collaboration in recent publications. The full publication of the *top 100 events* is in line with the Collaboration's commitment to sharing its data and results with the scientific community and to promote the exchange of knowledge between experiments.

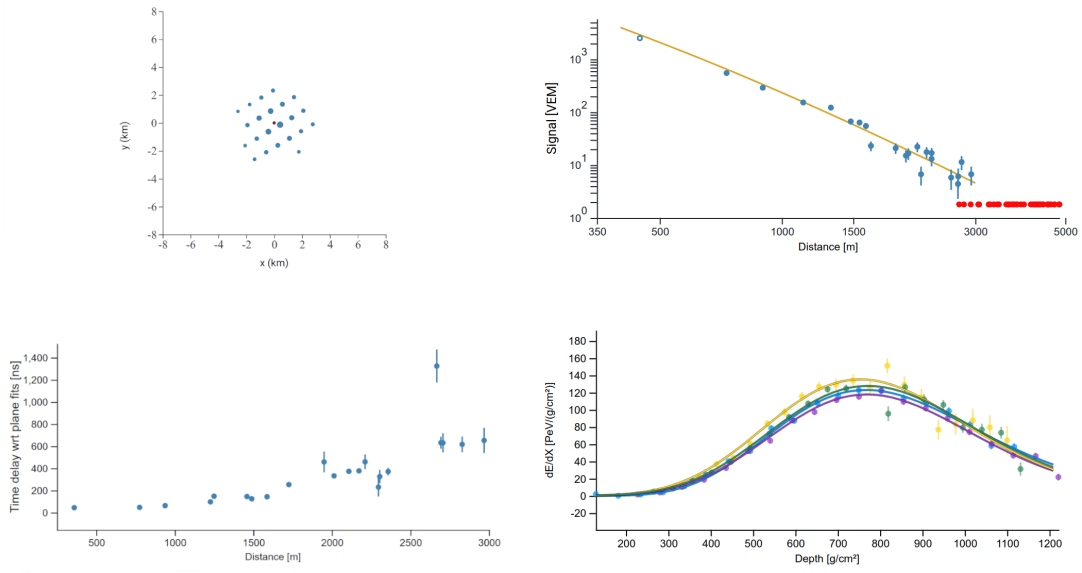

Figure 6: The highest-energy multi-eye hybrid event in the UHECR catalog [10], PAO100815 (id 102266222400): footprint with respect to the shower plane (top left panel); lateral distribution of the signals as a function of the distance from the shower axis (top right panel); time delays of the signals with respect to a fit with a plane shower front (bottom left panel) and reconstructed energy deposited in the atmosphere (bottom right panel).

4. Impact and use of our data

Open Data offer the basis for sharing knowledge with the scientific community and for developing diverse activities dedicated to the general public and high-school and higher-level students focused on learning physics and enjoying programming and data analysis.

A handful of scientific papers using the Auger Open Data have appeared in journals or on arXiv (https://arxiv.org) so far, but the use of the released open data is tracked directly via the Zenodo link (https://zenodo.org) and with Matomo tools (https://matomo.org). Since its first publication in 2021, the total number of visits has been more than 60 000 worldwide, while downloads of cosmic-ray data samples have been more than 4 000.

The data have been exploited in outreach events such as the International Cosmic Day, organized by DESY, in which students join the researchers and become scientists for one day. They have been used by teams of students in the IPPOG International Masterclasses program involving more than ten thousand 15- to 19-year-old students from 60 countries and taking place in about 225 universities or research centers [12]. The vast outreach program of the Pierre Auger Collaboration, including local and global initiatives with on-site visitors and students connected from many countries worldwide and an active presence on social networks and in the media, has generated a larger impact reaching tens of thousands of participants. For a comprehensive review of the outreach activities at the Pierre Auger Observatory, please refer also to [13, 14].

5. Perspectives

In June 2023, the Pierre Auger Collaboration Board approved the increase of the fraction of released cosmic-ray data to 30%. The Collaboration members are convinced that this will further boost the interest in using the Observatory data. The Observatory has completed its first phase of data taking. It has recently been upgraded with additional detectors, such as surface detector scintillators, underground muon detectors, radio antennas, and with upgraded electronics added to each surface detector station [15]. Future data will be easily integrated into this framework to produce Phase II open data, to the release of which the Auger Collaboration will undoubtedly maintain its dedication and commitment.

References

[1] Pierre Auger Collaboration [A. Aab, et al.], Nucl. Instrum. Meth. A **798** (2015) 172-213

[2] Wilkinson, M., et al., Sci Data **3** (2016) 160018; and *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*

[3] *Data open-access policy of the Pierre Auger Observatory*

[4] The Pierre Auger Observatory Open Data Portal https://opendata.auger.org/

[5] Pierre Auger Observatory Open Data, Zenodo, https://doi.org/10.5281/zenodo.4487612

[6] V. Scherini for the Pierre Auger Collaboration, PoS 37th ICRC (2021) 1386

[7] P. L. Ghia for the Pierre Auger Collaboration, PoS 38th ICRC (2023) 1616

[8] Pierre Auger Collaboration [A. Abdul Halim, et al.], Eur. Phys. J. C **85** (2025) 70

[9] Pierre Auger Collaboration [A. Abdul Halim, et al.], arXiv:2309.16294 [astro-ph]

[10] Pierre Auger Collaboration [A. Abdul Halim, et al.], Astrophys. J. Suppl. **264**, 2 (2023) 50

[11] Pierre Auger Collaboration [P. Abreu, et al.], ApJ **935** (2022) 170

[12] R. Sarmento for the Pierre Auger Collaboration, PoS 38th ICRC (2023) 1611

[13] K.S. Caballero-Mora for the Pierre Auger Collaboration, PoS 37th ICRC (2021) 1374

[14] B. Garcia for the Pierre Auger Collaboration, at this conference PoS 7th International Symposium on Ultra High Energy Cosmic Rays (UHECR 2024)

[15] Pierre Auger Collaboration [A. Aab, et al.], arXiv:1604.03637 [astro-ph]

PoS(UHECR2024)114