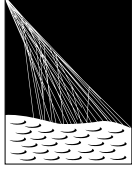


The Pierre Auger Observatory Open Data



PIERRE
AUGER
OBSERVATORY

A. Abdul Halim¹³, P. Abreu⁷⁰, M. Aglietta^{53,51}, I. Allekotte¹, K. Almeida Cheminant^{78,77}, A. Almela^{7,12}, R. Aloisio^{44,45}, J. Alvarez-Muñiz⁷⁶, A. Ambrosone⁴⁴, J. Ammerman Yebra⁷⁶, G.A. Anastasi^{57,46}, L. Anchordoqui⁸³, B. Andrada⁷, L. Andrade Dourado^{44,45}, S. Andringa⁷⁰, L. Apollonio^{58,48}, C. Aramo⁴⁹, P.R. Araújo Ferreira⁴¹, E. Arnone^{62,51}, J.C. Arteaga Velázquez⁶⁶, P. Assis⁷⁰, G. Avila¹¹, E. Avocone^{56,45}, A. Bakalova³¹, F. Barbato^{44,45}, A. Bartz Mocellin⁸², J.A. Bellido¹³, C. Berat³⁵, M.E. Bertaina^{62,51}, X. Bertou¹, M. Bianciotto^{62,51}, P.L. Biermann^a, V. Binet⁵, K. Bismark^{38,7}, T. Bister^{77,78}, J. Biteau^{36,i}, J. Blazek³¹, C. Bleve³⁵, J. Blümer⁴⁰, M. Boháčová³¹, D. Boncioli^{56,45}, C. Bonifazi⁸, L. Bonneau Arbeletche²², N. Borodai⁶⁸, J. Brack^f, P.G. Brichetto Orcherá⁷, F.L. Briechele⁴¹, A. Bueno⁷⁵, S. Buitink¹⁵, M. Buscemi^{46,57}, M. Büsken^{38,7}, A. Bwembya^{77,78}, K.S. Caballero-Mora⁶⁵, S. Cabana-Freire⁷⁶, L. Caccianiga^{58,48}, F. Campuzano⁶, R. Caruso^{57,46}, A. Castellina^{53,51}, F. Catalani¹⁹, G. Cataldi⁴⁷, L. Cazon⁷⁶, M. Cerda¹⁰, B. Čermáková⁴⁰, A. Cermenati^{44,45}, J.A. Chinellato²², J. Chudoba³¹, L. Chytka³², R.W. Clay¹³, A.C. Cobos Cerutti⁶, R. Colalillo^{59,49}, R. Conceição⁷⁰, A. Condorelli³⁶, G. Consolati^{48,54}, M. Conte^{55,47}, F. Convenga^{56,45}, D. Correia dos Santos²⁷, P.J. Costa⁷⁰, C.E. Covault⁸¹, M. Cristinziani⁴³, C.S. Cruz Sanchez³, S. Dasso^{4,2}, K. Daumiller⁴⁰, B.R. Dawson¹³, R.M. de Almeida²⁷, B. de Errico²⁷, J. de Jesús^{7,40}, S.J. de Jong^{77,78}, J.R.T. de Mello Neto²⁷, I. De Mitri^{44,45}, J. de Oliveira¹⁸, D. de Oliveira Franco⁴⁷, F. de Palma^{55,47}, V. de Souza²⁰, E. De Vito^{55,47}, A. Del Popolo^{57,46}, O. Deligny³³, N. Denner³¹, L. Deval^{40,7}, A. di Matteo⁵¹, C. Dobrigkeit²², J.C. D'Olivo⁶⁷, L.M. Domingues Mendes^{16,70}, Q. Dorosti⁴³, J.C. dos Anjos¹⁶, R.C. dos Anjos²⁶, J. Ebr³¹, F. Ellwanger⁴⁰, M. Emam^{77,78}, R. Engel^{38,40}, I. Epicoco^{55,47}, M. Erdmann⁴¹, A. Etchegoyen^{7,12}, C. Evoli^{44,45}, H. Falcke^{77,79,78}, G. Farrar⁸⁵, A.C. Fauth²², T. Fehler⁴³, F. Feldbusch³⁹, A. Fernandes⁷⁰, B. Fick⁸⁴, J.M. Figueira⁷, P. Filip^{38,7}, A. Filipčić^{74,73}, T. Fitoussi⁴⁰, B. Flagg⁸⁷, T. Fodran⁷⁷, M. Freitas⁷⁰, T. Fujii^{86,h}, A. Fuster^{7,12}, C. Galea⁷⁷, B. García⁶, C. Gaudu³⁷, P.L. Ghia³³, U. Giaccari⁴⁷, F. Gobbi¹⁰, F. Gollan⁷, G. Golup¹, M. Gómez Berisso¹, P.F. Gómez Vitale¹¹, J.P. Gongora¹¹, J.M. González¹, N. González⁷, D. Góra⁶⁸, A. Gorgi^{53,51}, M. Gottowik⁴⁰, F. Guarino^{59,49}, G.P. Guedes²³, E. Guido⁴³, L. Gülzow⁴⁰, S. Hahn³⁸, P. Hamal³¹, M.R. Hampel⁷, P. Hansen³, V.M. Harvey¹³, A. Haungs⁴⁰, T. Hebbeker⁴¹, C. Hojvat^d, J.R. Hörandel^{77,78}, P. Horvath³², M. Hrabovský³², T. Huege^{40,15}, A. Insolia^{57,46}, P.G. Isar⁷², P. Janecek³¹, V. Jilek³¹, J. Jurysek³¹, K.-H. Kampert³⁷, B. Keilhauer⁴⁰, A. Khakurdikar⁷⁷, V.V. Kizakke Covilakam^{7,40}, H.O. Klages⁴⁰, M. Kleifges³⁹, F. Knapp³⁸, J. Köhler⁴⁰, F. Krieger⁴¹, M. Kubatova³¹, N. Kunka³⁹, B.L. Lago¹⁷, N. Langner⁴¹, M.A. Leigui de Oliveira²⁵, Y. Lema-Capeans⁷⁶, A. Letessier-Selvon³⁴, I. Lhenry-Yvon³³, L. Lopes⁷⁰, J.P. Lundquist⁷³, A. Machado Payeras²², D. Mandat³¹, B.C. Manning¹³, P. Mantsch^d, F.M. Mariani^{58,48}, A.G. Mariazzi³, I.C. Mariş¹⁴, G. Marsella^{60,46}, D. Martello^{55,47}, S. Martinelli^{40,7}, O. Martínez Bravo⁶³, M.A. Martins⁷⁶, H.-J. Mathes⁴⁰, J. Matthews⁸, G. Matthiae^{61,50}, E. Mayotte⁸², S. Mayotte⁸², P.O. Mazur^d, G. Medina-Tanco⁶⁷, J. Meinert³⁷, D. Melo⁷, A. Menshikov³⁹, C. Merx⁴⁰, S. Michal³¹, M.I. Micheletti⁵, L. Miramonti^{58,48}, M. Mogarkar⁶⁸, S. Mollerach¹, F. Montanet³⁵, L. Morejon³⁷, K. Mulrey^{77,78}, R. Mussa⁵¹, W.M. Namasaka³⁷, S. Negi³¹, L. Nellen⁶⁷, K. Nguyen⁸⁴, G. Nicora⁹, M. Niechciol⁴³, D. Nitz⁸⁴, D. Nosek³⁰, A. Novikov⁸⁷, V. Novotny³⁰, L. Nožka³², A. Nucita^{55,47}, L.A. Núñez²⁹, C. Oliveira²⁰, M. Palatka³¹, J. Pallotta⁹, S. Panja³¹, G. Parente⁷⁶, T. Paulsen³⁷, J. Pawlowsky³⁷, M. Pech³¹, J. Pekala⁶⁸, R. Pelayo⁶⁴, V. Pelgrims¹⁴, L.A.S. Pereira²⁴, E.E. Pereira Martins^{38,7}, C. Pérez Bertolli^{7,40}, L. Perrone^{55,47}, S. Petrera^{44,45}, C. Petrucci⁵⁶, T. Pierog⁴⁰, M. Pimenta⁷⁰, M. Platino⁷, B. Pont⁷⁷, M. Pothast^{78,77}, M. Pourmohammad Shahvar^{60,46}, P. Privitera⁸⁶, M. Prouza³¹, S. Querschfeld³⁷, J. Rautenberg³⁷, D. Ravnani⁷, J.V. Reginatto Akim²², A. Reuzki⁴¹, J. Ridky³¹, F. Riehn⁷⁶, M. Risse⁴³, V. Rizi^{56,45}, E. Rodriguez^{7,40}, J. Rodriguez Rojo¹¹, M.J. Roncoroni⁷, S. Rossoni⁴², M. Roth⁴⁰, E. Roulet¹, A.C. Rovero⁴, A. Saftoiu⁷¹, M. Saharan⁷⁷, F. Salamida^{56,45}, H. Salazar⁶³, G. Salina⁵⁰, P. Sampathkumar⁴⁰, N. San Martín⁸², J.D. Sanabria Gomez²⁹, F. Sánchez⁷, E.M. Santos²¹, E. Santos³¹, F. Sarazin⁸², R. Sarmiento⁷⁰, R. Sato¹¹, P. Savina^{44,45}, C.M. Schäfer³⁸, V. Scherini^{55,47}, H. Schieler⁴⁰, M. Schimassek³³, M. Schimp³⁷, D. Schmidt⁴⁰, O. Scholten^{15,b}, H. Schoorlemmer^{77,78}, P. Schovánek³¹, F.G. Schröder^{87,40}, J. Schulte⁴¹, T. Schulz⁴⁰, S.J. Sciutto³, M. Scornavacche^{7,40}, A. Sedoski⁷, A. Segreto^{52,46},

S. Sehgal³⁷, S.U. Shivashankara⁷³, G. Sigl⁴², K. Simkova^{15,14}, F. Simon³⁹, R. Šmída⁸⁶, P. Sommers^e, R. Squartini¹⁰, M. Stadelmaier^{48,58,40}, S. Stanić⁷³, J. Stasielak⁶⁸, P. Stassi³⁵, S. Strähnz³⁸, M. Straub⁴¹, T. Suomijärvi³⁶, A.D. Supanitsky⁷, Z. Svozilikova³¹, Z. Szadkowski⁶⁹, F. Tairli¹³, A. Tapia²⁸, C. Taricco^{62,51}, C. Timmermans^{78,77}, O. Tkachenko³¹, P. Tobiska³¹, C.J. Todero Peixoto¹⁹, B. Tomé⁷⁰, Z. Torrès³⁵, A. Travaini¹⁰, P. Travnicek³¹, M. Tueros³, M. Unger⁴⁰, R. Uzeiroska³⁷, L. Vaclavek³², M. Vacula³², J.F. Valdés Galicia⁶⁷, L. Valore^{59,49}, E. Varela⁶³, V. Vašíčková³⁷, A. Vásquez-Ramírez²⁹, D. Veberić⁴⁰, I.D. Vergara Quispe³, S. Verpoest⁸⁷, V. Verzi⁵⁰, J. Vicha³¹, J. Vink⁸⁰, S. Vorobiov⁷³, C. Watanabe²⁷, A.A. Watson^c, A. Weindl⁴⁰, M. Weitz³⁷, L. Wiencke⁸², H. Wilczyński⁶⁸, D. Wittkowski³⁷, B. Wundheiler⁷, B. Yue³⁷, A. Yushkov³¹, O. Zapparrata¹⁴, E. Zas⁷⁶, D. Zavrtnik^{73,74}, M. Zavrtnik^{74,73}

The Pierre Auger Collaboration

————— • —————

- ¹ Centro Atómico Bariloche and Instituto Balseiro (CNEA-UNCuyo-CONICET), San Carlos de Bariloche, Argentina
- ² Departamento de Física and Departamento de Ciencias de la Atmósfera y los Océanos, FCEyN, Universidad de Buenos Aires and CONICET, Buenos Aires, Argentina
- ³ IFLP, Universidad Nacional de La Plata and CONICET, La Plata, Argentina
- ⁴ Instituto de Astronomía y Física del Espacio (IAFE, CONICET-UBA), Buenos Aires, Argentina
- ⁵ Instituto de Física de Rosario (IFIR) – CONICET/U.N.R. and Facultad de Ciencias Bioquímicas y Farmacéuticas U.N.R., Rosario, Argentina
- ⁶ Instituto de Tecnologías en Detección y Astropartículas (CNEA, CONICET, UNSAM), and Universidad Tecnológica Nacional – Facultad Regional Mendoza (CONICET/CNEA), Mendoza, Argentina
- ⁷ Instituto de Tecnologías en Detección y Astropartículas (CNEA, CONICET, UNSAM), Buenos Aires, Argentina
- ⁸ International Center of Advanced Studies and Instituto de Ciencias Físicas, ECyT-UNSAM and CONICET, Campus Miguelete – San Martín, Buenos Aires, Argentina
- ⁹ Laboratorio Atmósfera – Departamento de Investigaciones en Láseres y sus Aplicaciones – UNIDEF (CITEDEF-CONICET), Argentina
- ¹⁰ Observatorio Pierre Auger, Malargüe, Argentina
- ¹¹ Observatorio Pierre Auger and Comisión Nacional de Energía Atómica, Malargüe, Argentina
- ¹² Universidad Tecnológica Nacional – Facultad Regional Buenos Aires, Buenos Aires, Argentina
- ¹³ University of Adelaide, Adelaide, S.A., Australia
- ¹⁴ Université Libre de Bruxelles (ULB), Brussels, Belgium
- ¹⁵ Vrije Universiteit Brussels, Brussels, Belgium
- ¹⁶ Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro, RJ, Brazil
- ¹⁷ Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, Petropolis, Brazil
- ¹⁸ Instituto Federal de Educação, Ciência e Tecnologia do Rio de Janeiro (IFRJ), Brazil
- ¹⁹ Universidade de São Paulo, Escola de Engenharia de Lorena, Lorena, SP, Brazil
- ²⁰ Universidade de São Paulo, Instituto de Física de São Carlos, São Carlos, SP, Brazil
- ²¹ Universidade de São Paulo, Instituto de Física, São Paulo, SP, Brazil
- ²² Universidade Estadual de Campinas (UNICAMP), IFGW, Campinas, SP, Brazil
- ²³ Universidade Estadual de Feira de Santana, Feira de Santana, Brazil
- ²⁴ Universidade Federal de Campina Grande, Centro de Ciências e Tecnologia, Campina Grande, Brazil
- ²⁵ Universidade Federal do ABC, Santo André, SP, Brazil
- ²⁶ Universidade Federal do Paraná, Setor Palotina, Palotina, Brazil
- ²⁷ Universidade Federal do Rio de Janeiro, Instituto de Física, Rio de Janeiro, RJ, Brazil
- ²⁸ Universidad de Medellín, Medellín, Colombia

- 29 Universidad Industrial de Santander, Bucaramanga, Colombia
- 30 Charles University, Faculty of Mathematics and Physics, Institute of Particle and Nuclear Physics, Prague, Czech Republic
- 31 Institute of Physics of the Czech Academy of Sciences, Prague, Czech Republic
- 32 Palacky University, Olomouc, Czech Republic
- 33 CNRS/IN2P3, IJCLab, Université Paris-Saclay, Orsay, France
- 34 Laboratoire de Physique Nucléaire et de Hautes Energies (LPNHE), Sorbonne Université, Université de Paris, CNRS-IN2P3, Paris, France
- 35 Univ. Grenoble Alpes, CNRS, Grenoble Institute of Engineering Univ. Grenoble Alpes, LPSC-IN2P3, 38000 Grenoble, France
- 36 Université Paris-Saclay, CNRS/IN2P3, IJCLab, Orsay, France
- 37 Bergische Universität Wuppertal, Department of Physics, Wuppertal, Germany
- 38 Karlsruhe Institute of Technology (KIT), Institute for Experimental Particle Physics, Karlsruhe, Germany
- 39 Karlsruhe Institute of Technology (KIT), Institut für Prozessdatenverarbeitung und Elektronik, Karlsruhe, Germany
- 40 Karlsruhe Institute of Technology (KIT), Institute for Astroparticle Physics, Karlsruhe, Germany
- 41 RWTH Aachen University, III. Physikalisches Institut A, Aachen, Germany
- 42 Universität Hamburg, II. Institut für Theoretische Physik, Hamburg, Germany
- 43 Universität Siegen, Department Physik – Experimentelle Teilchenphysik, Siegen, Germany
- 44 Gran Sasso Science Institute, L'Aquila, Italy
- 45 INFN Laboratori Nazionali del Gran Sasso, Assergi (L'Aquila), Italy
- 46 INFN, Sezione di Catania, Catania, Italy
- 47 INFN, Sezione di Lecce, Lecce, Italy
- 48 INFN, Sezione di Milano, Milano, Italy
- 49 INFN, Sezione di Napoli, Napoli, Italy
- 50 INFN, Sezione di Roma "Tor Vergata", Roma, Italy
- 51 INFN, Sezione di Torino, Torino, Italy
- 52 Istituto di Astrofisica Spaziale e Fisica Cosmica di Palermo (INAF), Palermo, Italy
- 53 Osservatorio Astrofisico di Torino (INAF), Torino, Italy
- 54 Politecnico di Milano, Dipartimento di Scienze e Tecnologie Aerospaziali, Milano, Italy
- 55 Università del Salento, Dipartimento di Matematica e Fisica "E. De Giorgi", Lecce, Italy
- 56 Università dell'Aquila, Dipartimento di Scienze Fisiche e Chimiche, L'Aquila, Italy
- 57 Università di Catania, Dipartimento di Fisica e Astronomia "Ettore Majorana", Catania, Italy
- 58 Università di Milano, Dipartimento di Fisica, Milano, Italy
- 59 Università di Napoli "Federico II", Dipartimento di Fisica "Ettore Pancini", Napoli, Italy
- 60 Università di Palermo, Dipartimento di Fisica e Chimica "E. Segrè", Palermo, Italy
- 61 Università di Roma "Tor Vergata", Dipartimento di Fisica, Roma, Italy
- 62 Università Torino, Dipartimento di Fisica, Torino, Italy
- 63 Benemérita Universidad Autónoma de Puebla, Puebla, México
- 64 Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas del Instituto Politécnico Nacional (UPIITA-IPN), México, D.F., México
- 65 Universidad Autónoma de Chiapas, Tuxtla Gutiérrez, Chiapas, México
- 66 Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, México
- 67 Universidad Nacional Autónoma de México, México, D.F., México
- 68 Institute of Nuclear Physics PAN, Krakow, Poland
- 69 University of Łódź, Faculty of High-Energy Astrophysics, Łódź, Poland
- 70 Laboratório de Instrumentação e Física Experimental de Partículas – LIP and Instituto Superior Técnico – IST, Universidade de Lisboa – UL, Lisboa, Portugal
- 71 "Horia Hulubei" National Institute for Physics and Nuclear Engineering, Bucharest-Magurele, Romania
- 72 Institute of Space Science, Bucharest-Magurele, Romania
- 73 Center for Astrophysics and Cosmology (CAC), University of Nova Gorica, Nova Gorica,

- Slovenia
- ⁷⁴ Experimental Particle Physics Department, J. Stefan Institute, Ljubljana, Slovenia
- ⁷⁵ Universidad de Granada and C.A.F.P.E., Granada, Spain
- ⁷⁶ Instituto Galego de Física de Altas Enerxías (IGFAE), Universidade de Santiago de Compostela, Santiago de Compostela, Spain
- ⁷⁷ IMAPP, Radboud University Nijmegen, Nijmegen, The Netherlands
- ⁷⁸ Nationaal Instituut voor Kernfysica en Hoge Energie Fysica (NIKHEF), Science Park, Amsterdam, The Netherlands
- ⁷⁹ Stichting Astronomisch Onderzoek in Nederland (ASTRON), Dwingeloo, The Netherlands
- ⁸⁰ Universiteit van Amsterdam, Faculty of Science, Amsterdam, The Netherlands
- ⁸¹ Case Western Reserve University, Cleveland, OH, USA
- ⁸² Colorado School of Mines, Golden, CO, USA
- ⁸³ Department of Physics and Astronomy, Lehman College, City University of New York, Bronx, NY, USA
- ⁸⁴ Michigan Technological University, Houghton, MI, USA
- ⁸⁵ New York University, New York, NY, USA
- ⁸⁶ University of Chicago, Enrico Fermi Institute, Chicago, IL, USA
- ⁸⁷ University of Delaware, Department of Physics and Astronomy, Bartol Research Institute, Newark, DE, USA
-
- ^a Max-Planck-Institut für Radioastronomie, Bonn, Germany
- ^b also at Kapteyn Institute, University of Groningen, Groningen, The Netherlands
- ^c School of Physics and Astronomy, University of Leeds, Leeds, United Kingdom
- ^d Fermi National Accelerator Laboratory, Fermilab, Batavia, IL, USA
- ^e Pennsylvania State University, University Park, PA, USA
- ^f Colorado State University, Fort Collins, CO, USA
- ^g Louisiana State University, Baton Rouge, LA, USA
- ^h now at Graduate School of Science, Osaka Metropolitan University, Osaka, Japan
- ⁱ Institut universitaire de France (IUF), France

Email: spokespersons@auger.org

Abstract:

The Pierre Auger Collaboration has embraced the concept of open access to their research data since its foundation, with the aim of giving access to the widest possible community. A gradual process of release began as early as 2007 when 1% of the cosmic-ray data was made public, along with 100% of the space-weather information. In February 2021, a portal was released containing 10% of cosmic-ray data collected by the Pierre Auger Observatory from 2004 to 2018, during the first phase of operation of the Observatory. The Open Data Portal includes detailed documentation about the detection and reconstruction procedures, analysis codes that can be easily used and modified and, additionally, visualization tools. Since then, the Portal has been updated and extended. In 2023, a catalog of the highest-energy cosmic-ray events examined in depth has been included. A specific section dedicated to educational use has been developed with the expectation that these data will be explored by a wide and diverse community, including professional and citizen scientists, and used for educational and outreach initiatives. This paper describes the context, the spirit, and the technical implementation of the release of data by the largest cosmic-ray detector ever built and anticipates its future developments.

1 Introduction

During almost 20 years of data taking, the Pierre Auger Observatory [1], the largest facility for the measurement of ultra-high energy cosmic rays (UHECR), has detected more than 20 000 cosmic-ray events per year with an energy above 2.5 EeV ($1 \text{ EeV} = 10^{18} \text{ eV}$) providing, with unprecedented statistics and precision, major breakthroughs in the field.

The Observatory is located on a high-altitude plain near Malargüe, Mendoza Province, Argentina, at a mean altitude of about 1400 m, corresponding to an atmospheric overburden of about 875 g cm^{-2} . The site lies between latitudes 35.0°S and 35.3°S and between longitudes 69.0°W and 69.4°W . The Observatory combines two detection techniques for the detection of extensive air showers (EAS). A Surface Detector array of 1600 water-Cherenkov detectors (WCD) in a 1500 m triangular grid, SD-1500, yielding an effective area of 3000 km^2 , provides lateral sampling of the EAS at the ground. A Fluorescence Detector (FD) comprising 24 telescopes grouped at four sites (eyes) overlooks the array and detects the UV fluorescence light emitted by the de-excitation of the nitrogen molecules previously excited by the charged particles from the EAS.

Lasers for atmospheric monitoring are located towards the center of the array, at the positions marked as CLF (central laser facility) and XLF (extreme laser facility), see Fig. 1, left panel. The layout includes the low energy extension array, SD-750, with a spacing of 750 m covering 24 km^2 , and three high elevation telescopes (HEAT) installed at the FD-Coihueco site, see Fig. 1, right bottom corner.

The official data-taking started on 1 January 2004 with the engineering and the pre-production arrays of 154 surface detector stations and two operating fluorescence sites [2]. The installation of the SD-1500 array was completed in June 2008. The operation of the Observatory has been stable since then and is divided into two phases. During Phase I, completed on 31 December 2021, the Observatory consisted of the original layout. In Phase II, the Observatory has been upgraded with additional detectors, such as surface detector scintillators, underground muon detectors, and radio antennas, as well as upgraded electronics added to each surface detector station [3].

The rich data harvested by the Collaboration covers different and complementary fields of research. The main goal is to reveal the nature and origin of ultra-high energy cosmic rays, which relies on measurements of the energy spectrum and mass composition and is complemented by extensive searches for anisotropy in the distribution of the arrival directions at both large and intermediate angular scales. Ultra-high energy cosmic rays offer the unique opportunity to investigate the nature of astrophysical sources and particle interactions in a kinematic and energy region well beyond that covered by current particle accelerators. The main results include measurements of the cosmic-ray arrival directions [4], energy spectrum [5], and mass composition [6] along with their interpretation [7], the possible presence of neutral primaries such as photons and neutrinos [8], and particle physics [9]. Moreover, its potential as a multi-messenger [10] and multi-disciplinary [11, 12] observatory has been proven.

The Collaboration has embraced the concept of open access to research data from its foundation. A gradual process of releasing data began as early as 2007 when the Observatory was almost completed. A public event browser with 1% of data from the surface detector was created, and the data have been updated every year for over 10 years [13]. Meant for educational purposes, that portal was the first step towards making data publicly available at a time when there was no data management plan in place, and only a few other astroparticle physics experiments were releasing data before the end of their activities.

The Open Data Portal [14] was set up in February 2021, towards the end of the first phase of operation of the Observatory. It contained 10% of the cosmic-ray data used for the analyses presented in 2019 at the 36th International Cosmic Ray Conference in Madison, Wisconsin, US, comprising around 25 000 events from both the surface and the fluorescence detectors. Since then, it has been regularly updated and expanded in both the quantity and diversity of data in compliance with the *Data open-access policy of the Pierre Auger Observatory* [15].

Different sets of data were added to the original sample, starting with data from atmospheric measurements, both preprocessed and raw data, and data acquired in the so called “scaler mode”, recording the low-threshold counting rates of the surface detector stations. In December 2022,

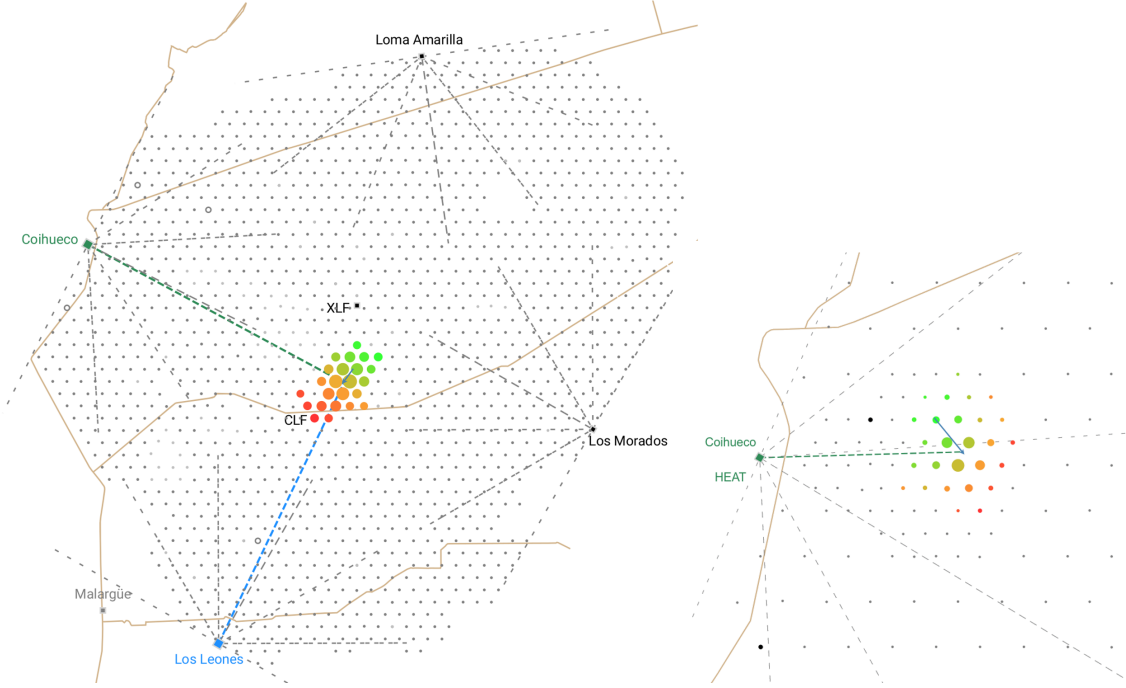


Figure 1: Map of the Observatory: left, surface detector stations (grey dots), position of the fluorescence detector eyes and field of view of the telescopes (squares, dashed lines) along with the footprint at the ground of one of the highest-energy multi-eye events in the released data sample (color scale from green to red reflects the arrival time of the shower particles at the ground). The positions of the atmospheric monitoring systems, XLF and CLF, are also shown. Right: footprint of an event detected by the SD-750 array in combination with the HEAT-Coihueco fluorescence telescopes, same color code

10% of cosmic-ray data above 60° was released, thus allowing to extend the aperture to a sky-coverage of 80%. A specific outreach section was included to engage the general public in cosmic-ray physics by supporting and facilitating the use of scientific data. In March 2024, 10% of cosmic-ray data detected by the low-energy extensions of the surface and fluorescence detectors was also released thus lowering the energy range down to 10^{17} eV.

It is worthwhile to emphasize that the data from the Observatory are the result of vast and long-term human and financial investment by the international community. The Collaboration is committed to their public release and provides accompanying software tools to offer to a wider community, including professional and citizen scientists, a unique opportunity to explore and analyse the data at various levels of complexity. This is in accordance with the Berlin declaration on Open Data and Open Access, inspired by the FAIR (Findable, Accessible, Interoperable, and Reusable) principles [16] for scientific data “as open as possible and as closed as necessary”. The Collaboration upholds the principle that open access to the data and associated software will, in the long term, allow the full scientific potential of the data to be realised.

This paper is conceived as follows: in Section 2 a description of the general structure and content of the Portal is given, while in Section 3, the several available datasets are described along with detailed tables with the files semantics provided in A. In Section 4 the event visualization is introduced, and, in Section 5, the tools for data analysis are presented. The catalog of the highest energy cosmic-ray events detected at the Observatory is described in Section 6. Section 7 is dedicated to the outreach page and the activities connected to Open Data. Finally, in Section 8, the experience of the Collaboration, after three years of monitoring the access and the use of the open data, is summarized, and the intentions for the future development of the open-access program are outlined.

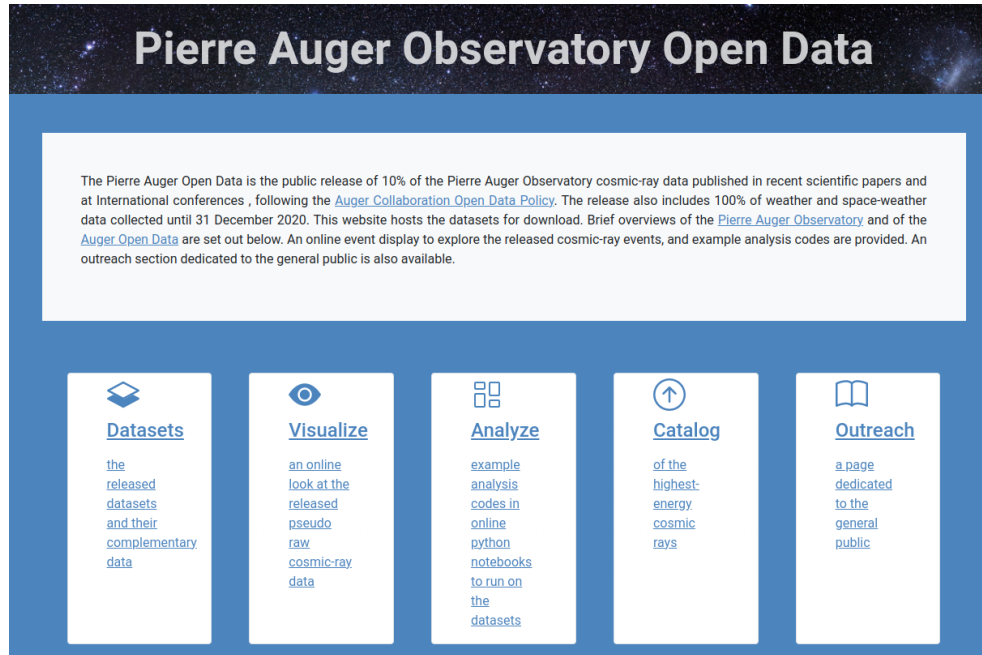


Figure 2: Screenshot of the main page of the Portal [14], with the links to the five different sections displayed. For more details see [17, 18]

2 The Portal

The Portal [14, 17, 18] is aimed at sharing a sensible fraction of the collected data and the associated knowledge with the scientific community. This can start from either raw experimental or simulated data through reconstructed events and datasets of higher level generated by analysis workflows, all the way up to data presented in scientific publications.

2.1 Portal Organization

An introductory main page provides links to the different cards via the navigation bar. This page also displays a comprehensive overview of the Observatory, regarding the main detectors and their working principles. Appropriate references are given for further investigation. Subsections explaining the details of the atmospheric monitoring program and the potential of the space-weather monitoring and observation of ionospheric phenomena are also provided. Finally, it includes the technical and copyright information. A screenshot of the main page is displayed in Fig. 2.

The *Datasets* card contains the list of releases, along with the available files and a detailed description of their structure and content. It also links to an exhaustive explanation of the data, and the conditions under which they were taken, reconstructed and selected, with links to the downloadable files (see Section 3).

The *Visualization* card provides a user-friendly interface for selecting and browsing each of the public events by specifying an event ID or a range of reconstructed variables, such as the energy or the zenith angle. Some exemplary events, such as the highest energy or multiplicity event, are pre-selected for inspection and available through a dedicated menu (see Section 4).

The *Analysis* card provides users with several tutorial codes to browse and analyse the data. Simple scripts for plotting histograms and graphs with the reconstructed variables and advanced routines based on the main Auger published results are available for download or can be run online (see Section 5).

The *UHECR Catalog* card implements a browser for the 100 highest-energy events recorded by the surface detector, along with the nine highest-energy hybrid events used for their calibration [19] (see Section 6).

The *Outreach* card, aimed at a wider audience, has been translated into several languages, providing a unique opportunity to share the excitement of cosmic-ray physics with the general public (see Section 7).

Finally, the Portal includes a *Contact* link giving the user additional support via answers to frequently asked questions about potential technical or content-related issues that might be raised. A dedicated e-mail address for contacting the Collaboration is given.

The datasets are released under the (CC BY-SA 4.0) International License [20], and they have a unique Digital Object Identifier (DOI) always pointing to the current version [21]. The user is requested to cite the general link or the specific version of the used data in any applications or publications. The DOIs of the specific released versions are detailed in Table 1.

Release Tag	Date	Content	Specific DOI
Release 3.0	March 20, 2024	10% cosmic-ray data, low energy sample	10.5281/zenodo.10488964
Release 2.0	Dec 22, 2022	10% cosmic-ray data, inclined sample ($60^\circ - 80^\circ$), Outreach section with different languages	10.5281/zenodo.6867688
Release 1.1	Oct 26, 2021	100% atmospheric data and scaler data	10.5281/zenodo.5588460
Release 1.0	Feb 15, 2021	10% cosmic-ray data, vertical sample ($0^\circ - 60^\circ$) and auxiliary files, ready-to-use event display, analysis examples	10.5281/zenodo.4487613

Table 1: List of specific releases on the Open Data Portal [14] with the corresponding DOIs

3 Datasets

Different types of data are provided via the Portal including cosmic-ray data, weather and space-weather data and other resources. Detailed explanations of the datasets, and the conditions under which data were collected and selected, are provided in the *Datasets* page together with the description of the available files and of all the data fields. Tables with the files semantics are given in A.

3.1 Cosmic-ray data

The cosmic-ray dataset comprises in total 81 121 showers 3348 of which are hybrid events, i.e. recorded in combination with the fluorescence detectors. The data are calibrated and partially cleaned to reduce the level of detailed understanding of the detectors required. The events are subjected to the reconstruction procedures implemented in the official software [22]. They are processed with the most up-to-date software at the time of the release and the software version is propagated in the files metadata. A set of selection criteria is applied to the detected cosmic-ray events in order to ensure an adequate sampling of the shower and the reliable performance of ground operation of the surface detector as well as of stable conditions of individual stations.

A total of 25 086 events measured with the SD-1500 have been selected. This set includes both vertical events, with a zenith angle less than 60° , and inclined events in the zenith interval

$60^\circ - 80^\circ$. Their reconstructed energies are above 2.5 EeV and 4 EeV respectively, to guarantee operation in an energy regime where detection with the surface array is fully efficient. Hybrid events are selected by imposing a set of criteria on the status of the hardware, reconstruction of shower geometry, shower profile, and atmospheric quality. Furthermore, specific fiducial volume cuts are applied for different analyses (energy spectrum, calibration, composition) to achieve uniform acceptance and minimize the uncertainties on the corresponding observables, with events being flagged accordingly.

The low energy sample comprises 54 481 events detected with the SD-750 array and reconstructed with a zenith angle less than 40° and energy above 0.1 EeV. Moreover, 197 low-energy hybrid events detected in combination with the HEAT-Coihueco telescopes and used for the energy calibration are also released. Further details of the currently released cosmic-ray samples are given in the specific sections of the *Datasets* page.

Each event is downloadable as pre-processed data in JavaScript Object Notation (JSON) files, structuring, in a compact format, the meta-data on the reconstruction along with blocks of data dedicated to each of the detectors involved in the event, see Table 2. The reconstructed parameters with their uncertainties and the list of the participating stations with the corresponding photomultiplier tubes (PMTs) traces are available. When an event is detected simultaneously with the fluorescence detector, the participating telescopes with the associated reconstructed energy-deposit profile and the signal recorded in the triggered pixels are provided. All the high-level reconstructed parameters from the surface and fluorescence detectors, such as energy, arrival direction, impact point at the ground, and depth of shower maximum ($X_{\max}$), are also given in a summary comma-separated (CSV) file. Details of the files semantics are given in Table 3.

Finally, auxiliary data files are also distributed, in particular for listing the positions of the surface detector stations and the angular field of view of the fluorescence detector pixels. The surface detector exposure and the parameters required to calculate the fluorescence detector acceptance for specific analyses are also provided. For details see Table 4.

3.2 Atmospheric data

The Observatory is a giant calorimeter, which includes the atmosphere as an important component. Therefore, a monitoring system has been set up at the detection site to measure local atmospheric parameters affecting the shower development, thus providing the knowledge required for the accurate reconstruction of observed air showers [23]. Changes in the atmospheric pressure lead to changes in the rates of recorded showers. At fixed pressure, if the temperature increases, the particles in the shower will spread out more as the distance travelled between each scattering rises. Variations in atmospheric properties also have significant effects on the rate of nitrogen fluorescence emission, as well as on the transmission of light.

The files produced by the atmospheric monitoring system include values of temperature, pressure, humidity, and wind speed recorded every five or ten minutes by five weather stations located at each fluorescence detector site and at the center of the array. In addition, the data set used to calculate the weather correction of the energy estimator derived from the surface detector, included in the standard reconstruction procedure, has also been released. The corresponding file, obtained by merging data from the weather stations, also contains the average value of the air density. Details about atmospheric data files are displayed in Table 5.

3.3 Space-weather data

Measurements of the background flux of secondary particles, produced mainly by low-energy cosmic rays (primary energies from 10 GeV to a few TeV) can be performed at the Observatory by exploiting scaler-mode data. The scaler mode is a particle counter mode, which is implemented for all the detectors of the surface array. These data are recorded for every station each second, reaching typical counting rates of 3 MHz. The temporal behavior of the number of counts is modulated by terrestrial and extraterrestrial phenomena and can thus be employed, for example, in studies of solar transient events like Forbush decreases and identification of modulations

related to the solar cycle [12]. From September 2005, the scaler mode has been used to count, in each of the 1600 detectors, the number of times the amplitude of the signals satisfies threshold conditions corresponding to an energy deposit of between 15 MeV and 100 MeV. The current data consists of more than 10^{15} signal counts detected until December 2020. Details about the scalers file content are given in Table 6.

4 Visualization

The characteristics of any shower in the dataset can be browsed via the *Visualization* page. It is possible to inspect the most interesting events from a menu, such as the highest-energy events, the highest multiplicity events, and multi-eye hybrid events. In addition, a browser allows to select events by their energy, zenith angle, number of triggered stations, and arrival time. Once an event is selected, its components can be browsed using different tabs. The event files can also be directly downloaded and further processed with the provided tutorial codes.

The *ground array view* tab displays a detailed map of the Observatory at the time of the event detection, showing the shower footprint at ground (color scale from green to red reflects the arrival time of the shower particles at the ground). The event shown in Fig. 1 with id 081847956000 is one of the highest energy multi-eye hybrid events in the released sample, with an energy of about 57 EeV and a zenith angle of 54° . It was detected on 03 July 2008 with 24 surface detector stations and simultaneously with two sites of the fluorescence detector. Further details of this exemplary event are shown in Fig. 3, Fig. 4, and Fig. 5.

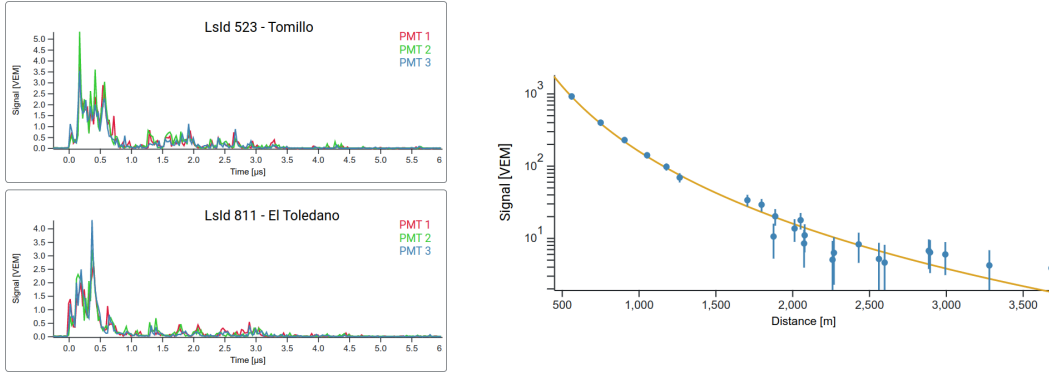


Figure 3: Visualization: the exemplary event, id 081847956000 (see text for details). Left panel: FADCs traces of the PMTs signals in two WCD stations participating in the event. Right: fall-off of the signal as a function of distance from the shower axis, the so-called lateral distribution function (LDF)

The flash analog-to-digital converters (FADCs) traces of the signals of the photomultiplier tubes of all triggered stations are displayed in the *SD traces* tab: the FADC traces from the PMTs of two WCD stations participating in the event are shown in Fig. 3, left panel. A fit of the signal fall-off as a function of distance from the shower impact point at the ground, the so-called lateral distribution function (LDF), provides the value of the energy estimator, the signal $S(r_{\text{opt}})$ at an optimal distance r_{opt} from the shower impact point on the ground in the plane perpendicular to the shower axis [24]. The LDF of the exemplary event recorded by the SD-1500 array is shown in Fig. 3, right panel. r_{opt} equals 1000 m for the SD-1500 array and 450 m for the SD-750 array, respectively.

For hybrid events, the *FD camera view* and *FD reconstruction* tabs contain information from the fluorescence telescopes. The sky view of the cameras and the reconstructed energy deposit as a

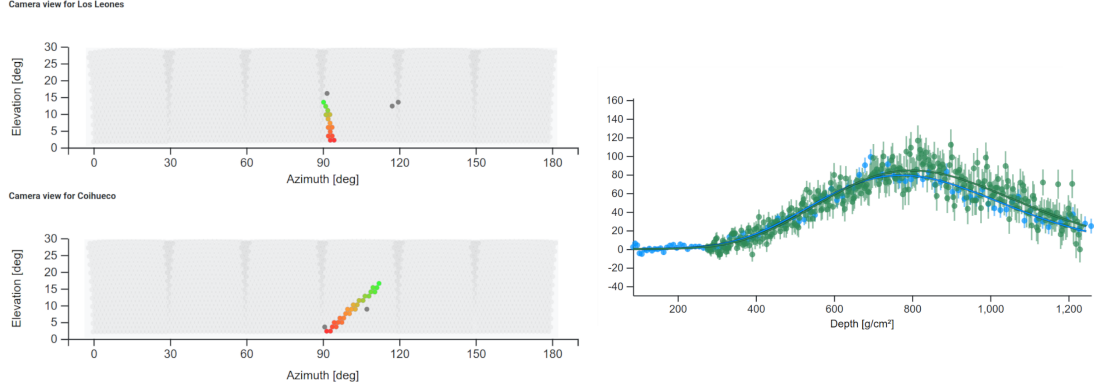


Figure 4: Visualization: the exemplary event, id 081847956000 (see text for details). Left panel: camera view of the fluorescence detector; the cosmic-ray shower is seen as a trace that moves along the pixels of the camera, from early (green) to late (red) pixels. Right panel: reconstructed energy deposit as a function of atmospheric depth as measured with the two telescopes participating in the event

function of atmospheric depth, the so-called longitudinal shower profile, are shown in Fig. 4, left and right panels, respectively. The energy deposited per unit depth in the atmosphere, dE/dX , increases at first, along with the multiplication of particles in the cascade, and then decreases as the rate of energy loss by ionisation starts to exceed that by radiative processes. This behavior gives rise to a universal profile shape [25], where the position of the maximum X_{max} depends mainly on the primary particle type (and its energy). The integration of the profile provides a calorimetric measurement of the total energy of the primary cosmic ray [26].

The *3D view* tab provides an interface to a 3-dimensional view of the events from different perspectives. An interactive and immersive view of the events and access to all reconstructed values and graphs are provided. A screenshot of the three-dimensional display for the exemplary event is shown in Fig. 5.

5 Analysis

The Open Data can be analyzed using Python Jupyter Notebooks [27]. Examples are provided in the Portal, together with a tutorial introducing the Python programming language and its use with the Open Data. These Notebooks are mostly designed to require only the core Python analysis packages and can be downloaded or run online in a web browser via the Kaggle platform [28].

Each cosmic-ray event can be processed via a simple Notebook exploring the content of the associated JSON file. Depending on the file structure and the data source, different plots can be produced, such as the PMTs signals of each surface detector station, the shower footprint at the ground, or the reconstructed profile of the energy deposited in the atmosphere.

Tutorials have been developed to show how to read the CSV summary files and the JSON files and produce plots using both pseudo-raw and higher-level data. The examples demonstrate how to produce histograms and plot the trend of variables as a function of time or energy, how to produce maps of the array and of arrival directions in the sky, and how to correlate the values of two variables. More advanced analysis codes are simplified re-implementations of parts of analyses published by the Collaboration, as detailed below.

The example analyses use the most-updated version of the Auger data sets and software, which may differ slightly from those used for associated publications because of improvements to the reconstruction and calibration procedures. The codes provided maintain the spirit of the

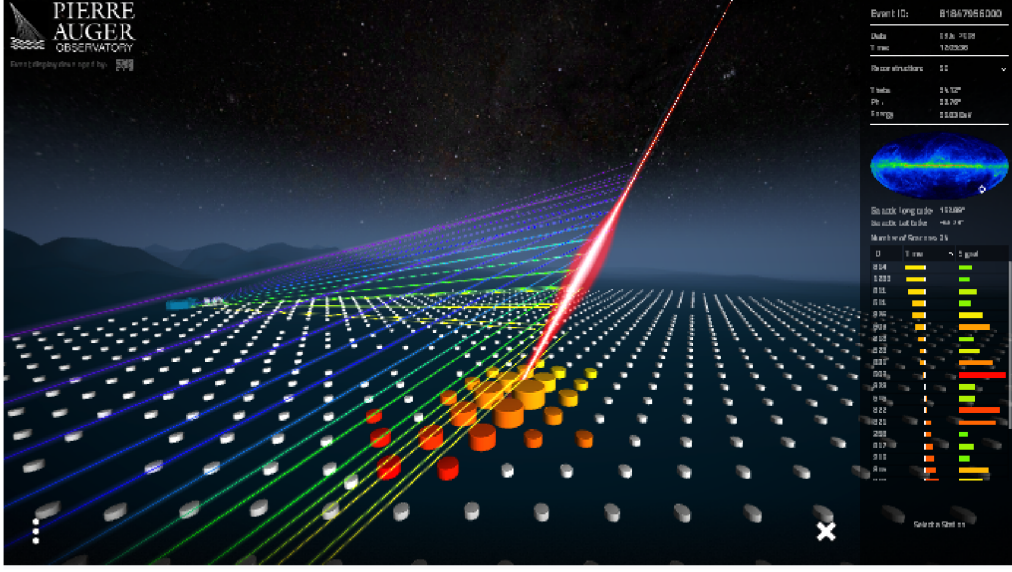


Figure 5: Visualization: a screenshot of the 3-dimensional display for the exemplary multi-eye hybrid event id 081847956000 (see text for details)

original analyses and provide insights as to how the results were obtained. They are simplified by omitting some more advanced analysis details that the user can find in the published papers. Even if the statistical significance is reduced with respect to what can be achieved with the full data set, the number of events is comparable to what was used in some of the first scientific publications by the Collaboration, see for example [29].

The energy calibration: The energy estimation for the vertical events recorded using the surface detector relies on the calibration of the energy estimator, $S(r_{\text{opt}})$. $S(r_{\text{opt}})$ is first corrected for the the zenith-dependent attenuation, deriving the signal that a shower would have produced when coming from the median of the zenith angle distribution (38° for the SD-1500 and 35° for the SD-750 sample). The calibration of the estimator is then performed by exploiting the calorimetric measurement of the energy made with the fluorescence detector for a sub-sample of high-quality hybrid events [5]. The correlation between the SD estimator and the energy measured with the fluorescence detector, E_{FD} , is plotted in Fig. 6, top left panel.

The energy spectrum: The estimation of the energy spectrum of cosmic rays detected with the surface detector is derived by counting the number of observed showers in differential energy bins and dividing it by the detector exposure. The bin size is constant in the logarithm of the energy, and the bin width corresponds approximately to the energy resolution. The energy threshold is fixed at 2.5 EeV for vertical events (zenith angle $< 60^\circ$), 4 EeV for inclined events ($60^\circ < \text{zenith angle} < 80^\circ$) and 1 EeV for the low energy sample, as this is the energy above which the surface detector acceptance becomes independent of the mass and energy of the primary cosmic ray. The energy spectra are shown in Fig. 6, top right panel. Further details are given in [5, 30, 31].

The depth of the shower maximum: The estimation of the atmospheric depth at which the deposited energy for a cosmic-ray shower as a function of atmospheric depth reaches its maximum relies on the reconstruction of the longitudinal profile of events measured by the fluorescence detector, and at least one coincident surface detector station (hybrid events) [6]. The X_{max} distributions in differential energy bins above 1 EeV for events with a zenith angle less than 75° are built, and the energy dependence of their mean and standard deviation is derived. These can be compared to those obtained from simulations of showers produced by proton and iron primaries. The rate of change of the mean X_{max} per decade of energy, the so-called elongation rate, is indicated in Fig. 6, bottom left panel.

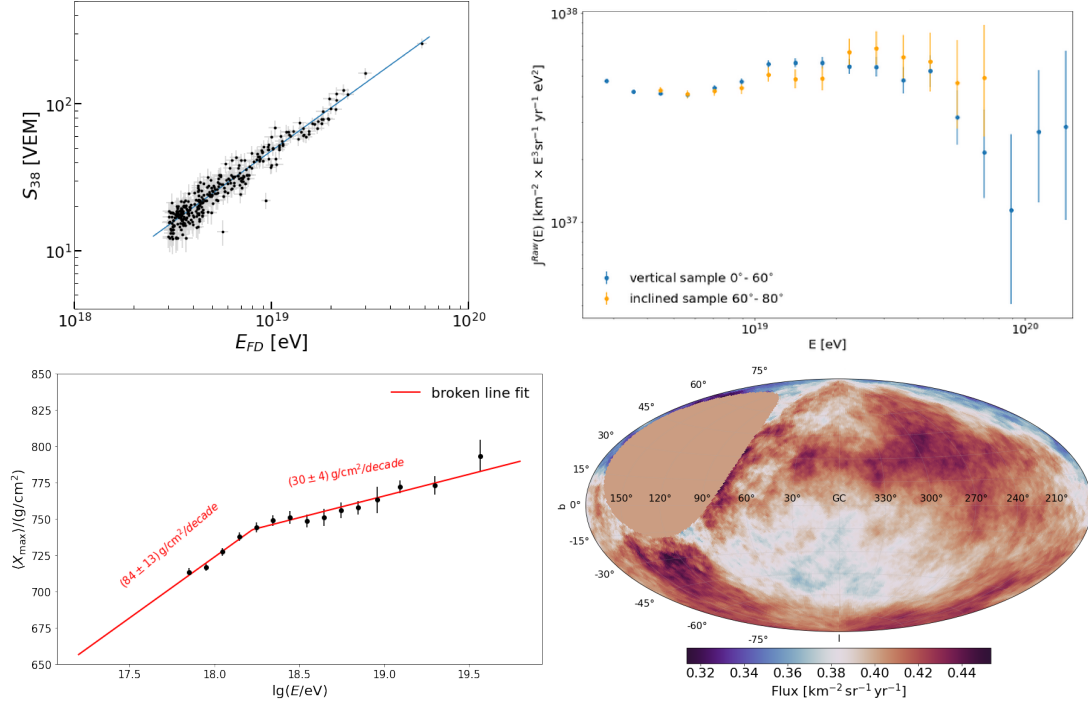


Figure 6: Analysis tab: exemplary plots resulting from the Python notebooks provided. Top left panel: energy calibration of the SD energy estimator (S_{38} for the SD-1500 array); top right panel: energy spectrum for vertical and inclined sample; bottom left panel: depth of shower maximum as a function of primary energy; bottom right panel: flux map for events with energy above 8 EeV

The measurement of the p-air cross-section: The proton-air cross section for particle production at a center-of-mass energy per nucleon of 57 TeV can be estimated by studying the shape of the distribution of $X_{\max}$. The attenuation length of primary cosmic ray protons in the atmosphere is reflected in the tail of the distribution at very high values of $X_{\max}$, which follows an exponential law [9].

The UHECR sky: The distribution of the reconstructed arrival directions of cosmic rays detected by the surface detector is studied to search for anisotropies at large angular scales [4]. The search is carried out by looking for non-uniformities in right ascension, as, for arrays that operate with close to 100% efficiency, the total exposure as a function of this angle is almost constant. A search for the first harmonic modulation in right ascension is performed by applying the Rayleigh formalism [32]. In Fig. 6, bottom right panel, the resulting smoothed flux map in Galactic coordinates for events with energies above 8 EeV is shown.

6 Catalog of the highest-energy cosmic-ray events

In 2023, the Pierre Auger Collaboration published a catalog of the 100 highest-energy cosmic-ray events [19] collected during Phase I of the data taking, along with the nine highest-energy hybrid events used for their calibration, demonstrating the quality of the data underlying the physics measurements carried out at the Observatory. In the paper, the instrumentation and the methods used to detect and reconstruct cosmic rays are also described.

The events are available for inspection and download from the *UHECR Catalog* page in the Portal. For each event, a summary of the reconstructed parameters and further details about the

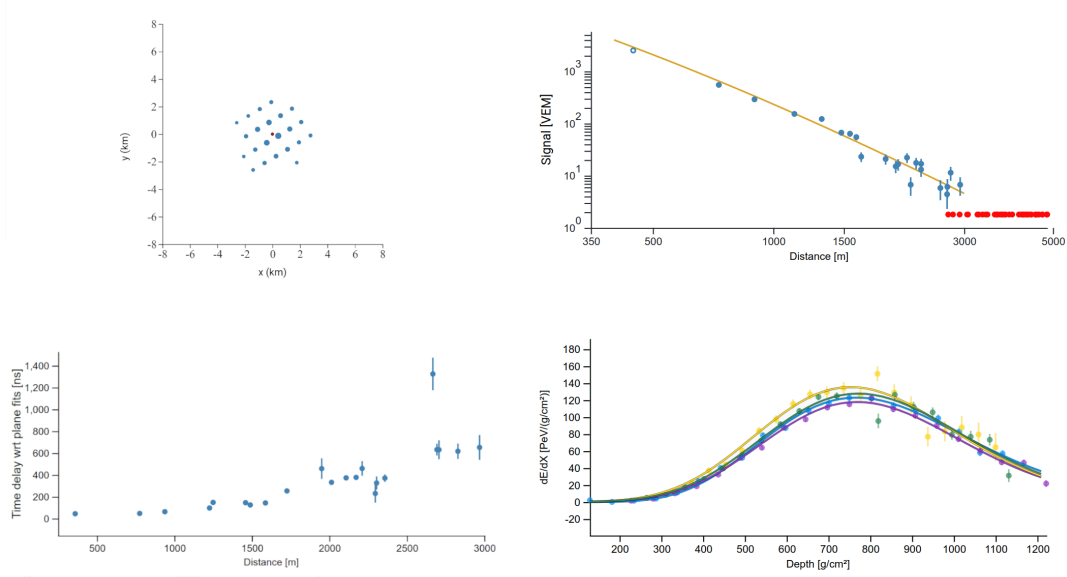


Figure 7: UHECR Catalog: the highest energy multi-eye hybrid event, PAO100815 (id 102266222400). Top left panel: footprint with respect to the shower plane; top right panel: lateral distribution of the signals as a function of the distance from the shower axis. Bottom left panel: time delays of the signals with respect to a fit with a plane shower front; bottom right panel: reconstructed energy deposited in the atmosphere

surface and fluorescence detector measurements are shown. The FADC traces of the photomultipliers of all triggered SD stations are also displayed in a dedicated tab.

In Fig. 7, the main characteristics of the highest energy multi-eye hybrid event in the catalog, PAO100815 (id 102266222400), that occurred on 15 August 2010, are shown. This event had a measured energy of 82 EeV and zenith angle of 53° , and it was hitting 22 surface detector stations and all four sites of the fluorescence detector. The footprint in a plane perpendicular to the shower arrival direction (top left panel) and the lateral distribution of the recorded signal as a function of the distance to the shower core in the log-log scale (top right panel) are also plotted. The user can view the time delays in nanoseconds with respect to a fit that assumes a plane shower front for the triggered stations (bottom left panel) and the reconstructed energy deposited in the atmosphere as measured by the fluorescence telescopes (bottom right panel).

7 Outreach

For many years the Pierre Auger Collaboration has carried out an extensive education and outreach program for disseminating its physics results, both locally and world-wide.

With the aim of facilitating its use by the general public, the Portal includes an *Outreach* page which is a “mini-framework” built in the same spirit as the research part, but with a simplified format. The same types of data are available in the outreach framework (cosmic rays, scalars, and meteorological data) and provided in summary CSV files. The page is implementing a multilingual “in-a-nutshell” description of the physics behind UHECR research and the techniques for observing extensive air showers at the ground. An overview of the Observatory detectors and their working principles is provided, with the help of some explanatory videos, along with a simplified description of the main achievements of the Collaboration. Exemplary tutorial and analysis tools have been developed and are linked to exercises dedicated to students, teachers,

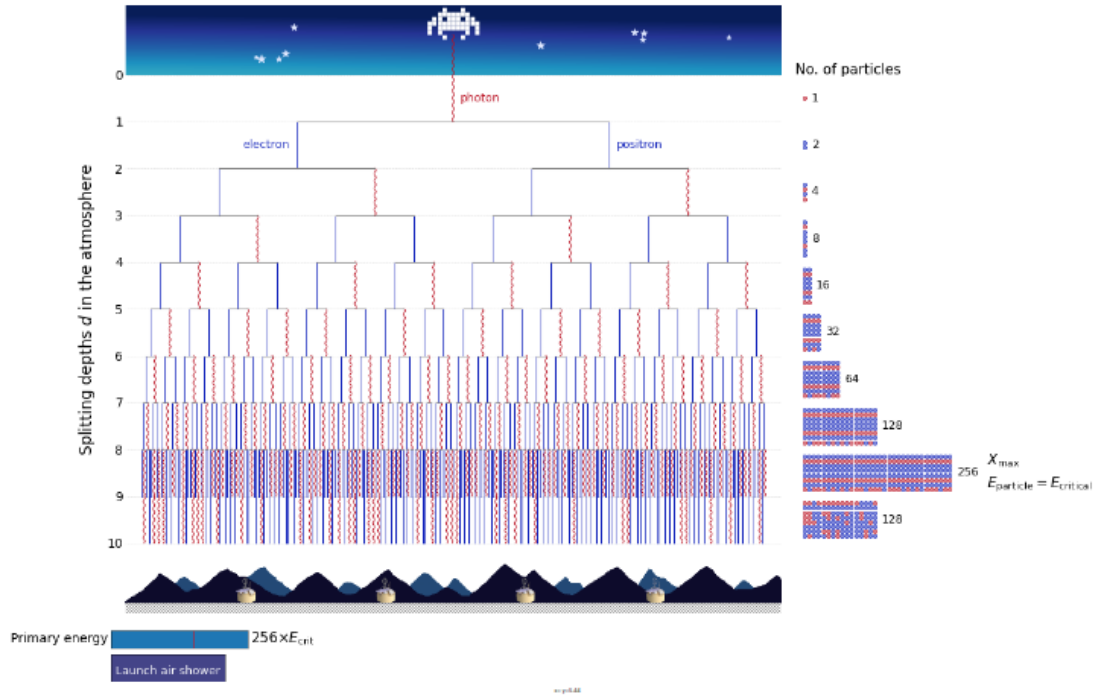


Figure 8: Toy model for the development of an electromagnetic cascade in the atmosphere

and citizen scientists in the general public, providing the necessary resources to understand the data and an invitation to use them for their own inquiry by developing original education and outreach activities. Some of the exercises are described below.

Exercise using surface detector data: The energy spectrum of cosmic-rays as detected by the surface detector is presented in this exercise, guiding the user to plot a histogram of the number of events detected in differential energy bins and normalizing it by the detector exposure. The exercise answers some of the questions the reader might have when approaching cosmic-ray physics: how the energy of a cosmic-ray primary particle relates to everyday life quantities, how rare are cosmic rays at these energies, and what is the energy carried by cosmic rays.

Exercise using hybrid data: This Notebook provides an example of how to carry out a simple analysis of the relative composition differences between cosmic rays coming from different parts of the sky. Data collected with the surface and the fluorescence detectors simultaneously, the so-called hybrid events, can be used to extract information related to primary cosmic-ray composition. The result is shown in a sky-map in Galactic coordinates.

Exercises using non cosmic-ray data: Two Notebooks are dedicated to non cosmic-ray data: atmospheric and space-weather data. Weather-station data are used to study the atmospheric conditions at the Observatory site. In particular, the exercise focuses on how to exploit weather stations data to calculate the value of air-density in different zones of the detection site. Data acquired via the so-called "scaler mode" can be used for space-weather science. In this Notebook the user can see how the event rate depends on the local weather conditions, such as pressure, temperature, and wind speed.

Exercise on the shower development: A simple toy-model for the description of the development of an extensive air shower is presented in this Notebook. An interactive example can be run to understand the process of particle multiplication occurring during shower development from the first interaction to reach billions of particles at the observation level. The sketch of the development of an electromagnetic cascade is shown in Fig. 8. A colour histogram of the longitudinal shower profile is plotted on the right hand-side.

8 Outlook

The Pierre Auger Collaboration has been committed, since the foundation of the Observatory, to opening the access to its data.

Data from the Observatory come from a variety of instruments and take many forms, starting from raw experimental data, through reconstructed events and datasets of higher level generated by analysis workflows all the way to data presented in scientific publications. Given the complexity and the diversity of the experimental data involved, the process of releasing them had to be started gradually.

The first step towards public data dates back to 2007 with a public event browser of 1% of cosmic-ray data. The Open Data Portal, firstly set up in February 2021 with 10% of the cosmic-ray events, has been updated and extended by including other types of data and a catalog of the highest-energy cosmic-ray events detected with the surface and the fluorescence detectors in the first phase of operation of the Observatory.

To facilitate the continued effort of the Portal maintenance, a task for Open Data has been created under the responsibility of the Project Management and in synergy with the related physics tasks. A standardized procedure has been designed to produce the simplified and portable format for Open Data from collaboration-internal, proprietary binary files. Code libraries are maintained within the Collaboration's software repository. Open Data are processed with the most up-to-date reconstruction software. Changes are propagated into updates of the released data, and identified in the metadata of the released files by the relative software version number. The validation of the data samples and the codes to be released is performed under the supervision of analysis and detector experts and certified by the related physics tasks. Finally, the accompanying documentation follows the standard internal review for publication.

Although only few scientific papers using the Open Data have appeared on journals or the ArXiv [33] so far, the visits to the Portal, tracked via Zenodo [34] and Matomo [35], number more than 40 000, while downloads of cosmic-ray samples number more than 3000 since its first publication in 2021.

Open Data offer the basis for developing diverse activities dedicated to high-school and higher-level students and to the general public, focused on learning physics and enjoying programming and data analysis. The Collaboration actively participates in various projects to engage the general public in cosmic-ray science and the physics of the Observatory by providing open-access to cosmic-ray data and tools for astro-particle physics dissemination. In this context, we quote the International Cosmic Day and the events carried out for the International Masterclasses program, organized within the IPPOG [36] consortium, involving thousands of high school students in different countries each year. For further details, please refer to [37].

In June 2023, the Collaboration Board approved the increase of the fraction of released cosmic-ray data to 30%, planned for 2024, to mark the 20th anniversary of the official start of data taking at the Observatory, that will be applied to the whole Phase I data of the Observatory. The members of the Collaboration are convinced that this will further boost the interest in and the use of the Observatory data. Future data from the upgraded Observatory, including new detectors, such as surface detector scintillators, underground muon detectors, and radio antennas, can be easily integrated into this framework to produce Phase II open data, to the release of which the Collaboration will undoubtedly maintain its commitment.

Acknowledgments

The successful installation, commissioning, and operation of the Pierre Auger Observatory would not have been possible without the strong commitment and effort from the technical and administrative staff in Malargüe. We are very grateful to the following agencies and organizations for financial support:

Argentina – Comisión Nacional de Energía Atómica; Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT); Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET); Gobierno de la Provincia de Mendoza; Municipalidad de Malargüe; NDM Holdings and Valle Las Leñas; in gratitude for their continuing cooperation over land access; Australia – the Australian Research Council; Belgium – Fonds de la Recherche Scientifique (FNRS); Research Foundation Flanders (FWO), Marie Curie Action of the European Union Grant No. 101107047; Brazil – Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); Financiadora de Estudos e Projetos (FINEP); Fundação de Amparo à Pesquisa do Estado de Rio de Janeiro (FAPERJ); São Paulo Research Foundation (FAPESP) Grants No. 2019/10151-2, No. 2010/07359-6 and No. 1999/05404-3; Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC); Czech Republic – GACR 24-13049S, CAS LQ100102401, MEYS LM2023032, CZ.02.1.01/0.0/0.0/16.013/0001402, CZ.02.1.01/0.0/0.0/18.046/0016010 and CZ.02.1.01/0.0/0.0/17.049/0008422 and CZ.02.01.01/00/22.008/0004632; France – Centre de Calcul IN2P3/CNRS; Centre National de la Recherche Scientifique (CNRS); Conseil Régional Ile-de-France; Département Physique Nucléaire et Corpusculaire (PNC-IN2P3/CNRS); Département Sciences de l’Univers (SDU-INSU/CNRS); Institut Lagrange de Paris (ILP) Grant No. LABEX ANR-10-LABX-63 within the Investissements d’Avenir Programme Grant No. ANR-11-IDEX-0004-02; Germany – Bundesministerium für Bildung und Forschung (BMBF); Deutsche Forschungsgemeinschaft (DFG); Finanzministerium Baden-Württemberg; Helmholtz Alliance for Astroparticle Physics (HAP); Helmholtz-Gemeinschaft Deutscher Forschungszentren (HGF); Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen; Ministerium für Wissenschaft, Forschung und Kunst des Landes Baden-Württemberg; Italy – Istituto Nazionale di Fisica Nucleare (INFN); Istituto Nazionale di Astrofisica (INAF); Ministero dell’Università e della Ricerca (MUR); CETEMPS Center of Excellence; Ministero degli Affari Esteri (MAE), ICSC Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by European Union NextGenerationEU, reference code CN.00000013; México – Consejo Nacional de Ciencia y Tecnología (CONACYT) No. 167733; Universidad Nacional Autónoma de México (UNAM); PAPIIT DGAPA-UNAM; The Netherlands – Ministry of Education, Culture and Science; Netherlands Organisation for Scientific Research (NWO); Dutch national e-infrastructure with the support of SURF Cooperative; Poland – Ministry of Education and Science, grants No. DIR/WK/2018/11 and 2022/WK/12; National Science Centre, grants No. 2016/22/M/ST9/00198, 2016/23/B/ST9/01635, 2020/39/B/ST9/01398, and 2022/45/B/ST9/02163; Portugal – Portuguese national funds and FEDER funds within Programa Operacional Factores de Competitividade through Fundação para a Ciência e a Tecnologia (COMPETE); Romania – Ministry of Research, Innovation and Digitization, CNCS-UEFISCDI, contract no. 30N/2023 under Romanian National Core Program LAPLAS VII, grant no. PN 23 21 01 02 and project number PN-III-P1-1.1-TE-2021-0924/TE57/2022, within PNCDI III; Slovenia – Slovenian Research Agency, grants P1-0031, P1-0385, I0-0033, N1-0111; Spain – Ministerio de Ciencia e Innovación/Agencia Estatal de Investigación (PID2019-105544GB-I00, PID2022-140510NB-I00 and RYC2019-027017-I), Xunta de Galicia (CIGUS Network of Research Centers, Consolidación 2021 GRC GI-2033, ED431C-2021/22 and ED431F-2022/15), Junta de Andalucía (SOMM17/6104/UGR and P18-FR-4314), and the European Union (Marie Skłodowska-Curie 101065027 and ERDF); USA – Department of Energy, Contracts No. DE-AC02-07CH11359, No. DE-FR02-04ER41300, No. DE-FG02-99ER41107 and No. DE-SC0011689; National Science Foundation, Grant No. 0450696; The Grainger Foundation; Marie Curie-IRSES/EPLANET; European Particle Physics Latin American Network; and UNESCO.

A Tables of file semantics

Section	Variable	Range, units and description
meta	type release format reconstruction software, version	Name of the release Version of the release: it defines the event sample Version of data format Software framework used for the event reconstruction and its version
info	id sdid gpstime date	Event identification number: YYDDSSSSSXX - YY: last 2 digits of year - DDD: day number between 1 and 366 - SSSS: second of the current day between 0 and 86399 - XX: order of the event at the current second. Time is expressed in UTC+12h, i.e., the day starting at noon Event number from data acquisition GPS time Date and time in ISO 8601 format
flags	sd1500 sd750 hdSpectrum hdCalib hdXmax multiEye	[0,1] 1: event is used in SD1500 array analysis [0,1] 1: event is used in SD750 array analysis [0,1] 1: event used for hybrid energy spectrum analysis [0,1] 1: event used for hybrid energy calibration analysis [0,1] 1: event used for hybrid Xmax analysis [0,1] 1: a multi-eye event
fdrec	id gpsnanotime hdSpectrumEye hdCalibEye hdXmaxEye theta, phi dtheta, dphi l, b ra, dec totalEnergy dtotalEnergy calEnergy dcalEnergy	[1-6] Indicates the FD site: '1': Los Leones '2': Los Morados '3': Loma Amarilla '4': Coihueco '5': HEAT '6': HEAT-Coihueco [ns] The GPS time of the event within its GPS second [0,1] 1: Eye used for the spectrum analysis [0,1] 1: Eye used for energy calibration analysis [0,1] 1: Eye used for Xmax analysis [deg] The zenith and azimuth angles [deg] Uncertainties in zenith and azimuth angles [deg] Galactic longitude and latitude of the event [deg] Right ascension and declination of the event [EeV] Total energy of the primary particle initiating the event [EeV] Uncertainty in the total energy of the event [EeV] Calorimetric energy of the event [EeV] Uncertainty in the calorimetric energy of the event

	pixelTime pixelCharge pixelStatus	[100 ns bin] Array of the times of the signal centroid in each pixel [n ph] Array of the light detected in each pixel (number of photons at telescope aperture) [0-4] Array that indicates the status of the pixel 0: Background 1: Triggered 2: Pulse 3: SDP (shower detector plane) 4: TimeFit
sdrec	gpsnanotime theta, dtheta phi, dphi energy denenergy l, b ra, dec x, y, z dx, dy easting, northing, altitude R dR s1000 ds1000 s450 ds450 s38 s35 n19 dn19 n68 dn68 gcorr wcorr beta,gamma chi2 ndf geochi2 geondf nbstat	[ns] GPS time [deg] Zenith angle and its uncertainty [deg] Azimuth angle and its uncertainty [EeV] Energy [EeV] Uncertainty in the energy [deg] Galactic longitude and latitude [deg] Right ascension and declination [m] Coordinate of the shower core (site coordinate system) [m] Uncertainty in the coordinates of the shower core (site c. s.) [m] Eastward, northward coordinates and altitude of the shower core (UTM coordinates system) [m] Radius of curvature of the shower [m] Uncertainty in the radius of curvature of the shower [VEM] Expected signal at 1000 m from the core (SD1500 array) [VEM] Uncertainty in S(1000) [VEM] Expected signal at 450 m from the core (SD750 array) [VEM] Uncertainty in S(450) [VEM] Signal produced at 1000 m by a shower with a zenith angle of 38° [VEM] Signal produced at 450 m by a shower with a zenith angle of 35° Energy estimator, N19, of a shower with a zenith angle > 60° Uncertainty in N19 N19, that a shower would have produced had it arrived at 68° Uncertainty in N68 [%] Geomagnetic correction to S(1000) [%] Weather correction to S(1000) Slope parameters of the fitted LDF Chi-square value of the LDF fit Number of degrees of freedom in the LDF fit Chi-square value of the geometric fit Number of degrees of freedom in the geometric fit Number of triggered stations used in reconstruction

	recstations	List of ids of the triggered stations used in reconstruction
stations	id name x,y,z t dt signalStartBin,signalStopBin signal dsignal sat isSelected spDistance dspDistance pmt1,pmt2,pmt3	Id of the station Name of the station [m] Coordinates of the station [ns] Start time of the signal [ns] Uncertainty in the start time FADC trace bins that indicate the start and stop of the signal [VEM] Integrated signal in the FADC traces [VEM] Uncertainty in the integrated signal [0-2] 0: high-gain and low-gain channels not saturated 1: high-gain channel saturated 2: high-gain and low-gain channels saturated [0-1] 1: the station is used in the reconstruction [m] Distance of the station to the core in the plane perpendicular to the shower axis (shower plane) [m] Uncertainty in spDistance [VEM] FADC traces from each photomultiplier. The length of each FADC trace is 768 bins. A bin corresponds to 25 ns

Table 2: Cosmic-ray dataset - content of the JSON file

Filename	Variable	Range, units and description
dataSummarySD1500.csv dataSummaryInclned.csv dataSummarySD750.csv	id sdid gpstime sd750 multiEye sd_gpsnanotime sd_theta, sd_phi sd_dtheta, sd_dphi sd_energy sd_denergy	Event identification number: YYD-DDSSSSXX - YY: last 2 digits of year - DDD: day number between 1 and 366 - SSSS: second of the current day between 0 and 86399 - XX: order of the event at the current second. Time is expressed in UTC+12h, i.e., the day starting at noon Event number from data acquisition GPS time of the eventad 1: event is used in SD1500 array analysis [0,1] 1: event is used in SD750 array analysis [0,1] 1: a multi-eye event [ns] The GPS time of the event within its GPS second [deg] The zenith and azimuth angles deg] Uncertainties in the zenith and azimuth angles [EeV] Total energy of the primary particle initiating the event [EeV] Uncertainty in the total energy of the event

sd_l, sd_b	[deg] Galactic longitude and latitude
sd_ra, sd_dec	[deg] Right ascension and declination
sd_x, sd_y sd_z	[m] Coordinates of the shower core (site coordinates system)
sd_dx, sd_dy	[m] Uncertainties in the x, y coordinates of the shower core
sd_easting, sd_northing	[m] Eastward and northward coordinates of the shower core (UTM c.s.)
sd_altitude	[m] Altitude of the shower core (UTM coordinates system)
sd_R	[m] Radius of curvature of the shower
sd_dR	[m] Uncertainty in the radius of curvature of the shower
sd_s1000	[VEM] Expected signal at 1000 m from the core
sd_ds1000	[VEM] Uncertainty in S(1000)
sd_s38	[VEM] Signal produced at 1000 m by a shower with a zenith angle of 38°
sd_gcorr	Geomagnetic correction to S(1000)
sd_wcorr	Weather correction to S(1000)
sd_beta, sd_gamma	Slope parameters of the fitted LDF
sd_chi2	Chi-square value of the LDF fit
sd_ndf	Number of degrees of freedom in the LDF fit
sd_geochi2	Chi-square value of the geometric fit
sd_geondf	Number of degrees of freedom in the geometric fit
sd_nbstat	Number of triggered stations used in reconstruction
fd_id	[1-6] Indicates the FD site
fd_gpsnanotime	[ns] The GPS time of the event within its GPS second
fd_hdSpectrumEye	[0,1] 1: Eye used for the spectrum analysis
fd_hdCalibEye	[0,1] 1: Eye used for energy calibration analysis
fd_hdXmaxEye	[0,1] 1: Eye used for Xmax analysis
fd_theta, fd_phi	[deg] The zenith and azimuth angles
fd_dtheta, fd_dphi	[deg] Uncertainties in zenith and azimuth angles
fd_l, fd_b	[deg] Galactic longitude and latitude of the event
fd_ra, fd_dec	[deg] Right ascension and declination of the event
fd_totalEnergy	[EeV] Total energy of the primary particle initiating the event
fd_dtotalEnergy	[EeV] Uncertainty in the total energy of the event
fd_calEnergy	[EeV] Calorimetric energy of the event
fd_dcalEnergy	[EeV] Uncertainty in the calorimetric energy of the event

	fd.xmax	[g/cm ²] Position of the maximum of the shower longitudinal development in the atmosphere
	fd.dxmax	[g/cm ²] Uncertainty in xmax
	fd.heightXmax	[m a.s.l.] Height of Xmax above sea level
	fd.distXmax	[m] Distance of Xmax to FD eye
	fd.dEdXmax	[PeV/(g/cm ²)] Maximum energy deposit
	fd.ddEdXmax	[PeV/(g/cm ²)] Uncertainty in the maximum energy deposit
	fd.x, fd.y, fd.z	[m] Coordinates of the shower core projected at ground level (site coordinates system)
	fd.dx, fd.dy	[m] Uncertainty in the coordinates of the shower core projected at ground level (site coordinates system)
	fd.easting, fd.northing	[m] Eastward and northward coordinates of the shower core projected at ground level (UTM coordinates system)
	fd.altitude	[m] Altitude of the shower core projected at ground level (UTM coordinates system)
	fd.cherenkovFraction	Fraction of detected light from Cherenkov emission
	fd.minViewAngle	[deg] Light emission angle from the shower towards the FD
	fd.uspL, fd.duspL	[g/cm ²] Universal shower profile shape parameter L and its uncertainty
	fd.uspR, fd.duspR	Universal shower profile shape parameter R and its uncertainty
	fd.hottestStationId	Id of the SD station with the highest recorded signal (hottest)
	fd.distSdpStation	[m] Distance of the hottest station to the shower detector plane
	fd.distAxisStation	[m] Distance of hottest station to the reconstructed shower axis
	sd.exposure	Value of the exposure (for the SD1500 or SD750 array) at the time of the event rescaled by the data release fraction

Table 3: Cosmic-ray dataset - content of the summary CSV files

File name	Variable	Description
<i>sdMap.csv</i>	id	Identification number of the station
	northing	[m] UTC coordinates: northing
	easting	[m] UTC coordinates: easting
	altitude	[m] UTC coordinates: altitude
	start	GPS time of the first event detected by the station
	stop	GPS time of the last event detected by the station
	sd1500	[0,1] 1: station is part of the SD1500 array
	sd750	[0,1] 1: station is part of the SD750 array

File name	Variable	Description
<i>weather.csv</i>	time temperature pressure density avgDensity2HoursBefore	[s] Unix time (seconds since 1st Jan 1970, w/o leap seconds) [°C] Air temperature [hPa] Barometric pressure [kg/m ³] Air density [kg/m ³] Value of air-density measured two hours earlier
<i>wsLosLeones.csv</i> , <i>wsLosMorados.csv</i> , <i>wsLomaAmarilla.csv</i> , <i>wsCoihueco.csv</i> , <i>wsCLF.csv</i>	time temperature humidity windSpeed pressure	[s] Unix time (seconds since 1st Jan 1970, w/o leap seconds) [°C] Air temperature [%] Relative humidity [km/h] Average wind speed [hPa] Barometric pressure

Table 5: Atmospheric dataset - content of the weather stations files

File name	Variable	Description
<i>scalers.csv</i>	time rateCorr arrayFraction rateUncorr pressure	[s] Unix time (seconds since 1st Jan 1970, w/o leap seconds) [counts/m ² /s] Corrected scaler rate [%] Fraction of array in operation [counts/s] Average detector scaler rate, uncorrected [hPa] barometric pressure

Table 6: Scaler dataset - content of the scaler mode file

References

- [1] Pierre Auger Collaboration [A. Aab, et al.], *The Pierre Auger Cosmic Ray Observatory*, *Nucl. Instrum. Meth. A* **798** (2015) 172-213
- [2] Pierre Auger Collaboration [J. Abraham et al.], *Properties and performance of the prototype instrument for the Pierre Auger Observatory*, *Nucl. Instrum. Meth. A* **523** (2004) 50
- [3] Pierre Auger Collaboration [A. Aab, et al.], *The Pierre Auger Observatory Upgrade - Preliminary Design Report*, *Report FERMILAB-DESIGN-2016-05*
- [4] Pierre Auger Collaboration [A. Aab, et al.], *Observation of a large-scale anisotropy in the arrival directions of cosmic rays above 8×10^{18} eV*, *Science* **357** (2017) 1266; Pierre Auger Collaboration [P. Abreu, et al.], *Arrival Directions of Cosmic Rays above 32 EeV from Phase One of the Pierre Auger Observatory*, *Astrophys. J.* **935** (2022) 170
- [5] Pierre Auger Collaboration [A. Aab, et al.], *Measurement of the cosmic-ray energy spectrum above 2.5×10^{18} eV using the Pierre Auger Observatory*, *Phys. Rev. D* **102** (2020) 062005; Pierre Auger Collaboration [A. Aab, et al.], *Features of the Energy Spectrum of Cosmic Rays above 2.5×10^{18} eV Using the Pierre Auger Observatory*, *Phys. Rev. Lett.* **125** (2020) 121106
- [6] Pierre Auger Collaboration [A. Aab, et al.], *Depth of maximum of air-shower profiles at the Pierre Auger Observatory. I. Measurements at energies above $10^{17.8}$ eV*, *Phys. Rev. D* **90** (2014) 122005

- [7] Pierre Auger Collaboration [A. Abdul Halim, et al.], *Constraining the sources of ultra-high-energy cosmic rays across and above the ankle with the spectrum and composition data measured at the Pierre Auger Observatory*, JCAP **05** (2023) 024
- [8] Pierre Auger Collaboration [P. Abreu, et al.], *Search for photons above 10^{19} eV with the surface detector of the Pierre Auger Observatory*, JCAP **05** (2023) 021; Pierre Auger Collaboration [A. Aab, et al.], *Probing the origin of ultra-high-energy cosmic rays with neutrinos in the EeV energy range using the Pierre Auger Observatory*, JCAP **10** (2019) 022
- [9] Pierre Auger Collaboration [P. Abreu, et al.], *Measurement of the Proton-Air Cross Section at $\sqrt{s} = 57$ TeV with the Pierre Auger Observatory*, Phys. Rev. Lett. **109** (2012) 062002
- [10] Pierre Auger Collaboration [A. Aab, et al.], *Multi-messenger physics with the Pierre Auger Observatory*, Front. Astron. Space Sci. **6** (2019) 24
- [11] Pierre Auger Collaboration [A. Aab, et al.], *A 3-Year Sample of Almost 1,600 Elves Recorded Above South America by the Pierre Auger Cosmic-Ray Observatory*, Earth and Space Science, **7** (2020) e2019EA000582; R. Colalillo for the Pierre Auger Collaboration, *Downward Terrestrial Gamma-ray Flashes at the Pierre Auger Observatory?*, PoS ICRC2021 395
- [12] Pierre Auger Collaboration [P. Abreu, et al.], *The Pierre Auger Observatory scaler mode for the study of solar activity modulation of galactic cosmic rays*, JINST, **6** (2011) P01003; M. Schimassek for the Pierre Auger Collaboration, *Analysis of Data from the Low-Energy Modes of the Surface Detector of the Pierre Auger Observatory*, PoS ICRC2019 1147
- [13] The Pierre Auger Observatory Public Event Browser <https://labdpr.org/ED/index.php>
- [14] The Pierre Auger Observatory Open Data Portal <https://opendata.auger.org/>
- [15] The Pierre Auger Collaboration, *Data open-access policy of the Pierre Auger Observatory*
- [16] Wilkinson, M., et al., *The FAIR Guiding Principles for scientific data management and stewardship*, Sci Data **3** (2016) 160018; and *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*
- [17] V. Scherini for the Pierre Auger Collaboration, *The 2021 Open-Data release by the Pierre Auger Collaboration*, PoS 37th ICRC (2021) 1386
- [18] P. L. Ghia for the Pierre Auger Collaboration, *Portals to data of the Pierre Auger Observatory*, PoS 38th ICRC (2023) 1616
- [19] Pierre Auger Collaboration [A. Abdul Halim, et al.], *A Catalog of the Highest-energy Cosmic Rays Recorded during Phase I of Operation of the Pierre Auger Observatory*, Astrophys. J. Suppl. **264**, 2 (2023) 50
- [20] Creative Common (CC BY-SA 4.0) International License <https://creativecommons.org/licenses/by-sa/4.0>
- [21] Pierre Auger Observatory Open Data, Zenodo, <https://doi.org/10.5281/zenodo.4487612>
- [22] Pierre Auger Collaboration [J. Abraham et al.], *The Offline Software Framework of the Pierre Auger Observatory*, Nucl. Instrum. Meth. A **580** (2007) 1485-1496
- [23] Pierre Auger Collaboration [A. Aab, et al.], *Impact of atmospheric effects on the energy reconstruction of air showers observed by the surface detectors of the Pierre Auger Observatory*, JINST **12**, 2 (2017) P022006
- [24] Pierre Auger Collaboration [A. Aab, et al.], *Reconstruction of events recorded with the surface detector of the Pierre Auger Observatory*, JINST **15** (2020) P10021

- [25] Pierre Auger Collaboration [A. Aab, et al.], *Measurement of the average shape of longitudinal profiles of cosmic-ray air showers at the Pierre Auger Observatory*, JCAP **03** (2019) 018
- [26] Pierre Auger Collaboration [A. Aab et al.], *Data-driven estimation of the invisible energy of cosmic ray showers with the Pierre Auger Observatory*, Phys. Rev. D **100** (2019) 082003
- [27] Jupyter Notebook Interface, <https://jupyter.org/>
- [28] Kaggle Community Interface, <https://www.kaggle.com/>
- [29] Pierre Auger Collaboration [J. Abraham et al.], *Observation of the Suppression of the Flux of Cosmic Rays above 4×10^{19} eV*, Phys. Rev. Lett. **101** (2008) 061101
- [30] Pierre Auger Collaboration [A. Aab, et al.], *Measurement of the cosmic ray spectrum above 4×10^{18} eV using inclined events detected with the Pierre Auger Observatory*, JCAP **08** (2015) 49
- [31] Pierre Auger Collaboration [P. Abreu, et al.], *The energy spectrum of cosmic rays beyond the turn-down around 10^{17} eV as measured with the surface detector of the Pierre Auger Observatory*, Eur. Phys. Journ. C **81** (2021) 966
- [32] J. Linsley, *Fluctuation Effects on Directional Data*, Phys. Rev. Lett. **34** (1975) 1530
- [33] ArXiv e-Print archive, <https://arxiv.org/>
- [34] Zenodo, <https://zenodo.org/>
- [35] Matomo Analytics, <https://matomo.org>
- [36] International Particle Physics Outreach Group, <https://ippog.org/>
- [37] R. Sarmiento for the Pierre Auger Collaboration, *International Masterclasses as part of the Pierre Auger Observatory program of Outreach and Education*, PoS 38th ICRC (2023) 1611