

ISAPP 2025 Lecce - 9-20 June 2025

Open Data in Science: the case of the Pierre Auger Observatory

Viviana Scherini Università del Salento and INFN Lecce



Outline

Open Data in science: general thoughts

Exemplary case: the Pierre Auger Collaboration Open Data

- motivations
- challenges
- organization

Design and implementation of open data policy

- history & current status \rightarrow portal description, events & analysis
- impact of open data \rightarrow science & outreach
- future steps \rightarrow increase fraction, phase II data

Open data (FAIR) principle

Open science is an approach to research based on open cooperative work that emphasizes the sharing of knowledge, results and tools as early and widely as possible

FAIR (Findable, Accessible, Interoperable, and Reusable)

- easy to be openly used, reused, retained and redistributed by anyone
- available in a timely and user-friendly, human- and machine-readable
- supported by regular curation and maintenance

Open Research

- digital and analog data, both **raw** and **processed**
- **metadata** (data about data) numerical scores, textual records, images
- analysis codes and workflows

Motivations & Challenges

Goal of maximizing scientific potential of collected data

- + **share** with scientific community: **multi-messenger effort**
- + verify results and possibly conduct new analyses
- + distribute knowledge to a broader audience (tax payers)

Challenges

- complex data structures \rightarrow huge effort needed to simplify and make usable
- need to define data ownership and proper credit to the Collaboration
- make the process effective on the long term by balancing with data collection and maintenance activities

Open data in astroparticle experiments

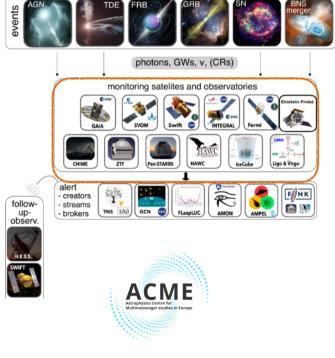
- large collaborations
- diverse and collective duties
 - detector design & construction
 - operation & maintenance
 - simulation & data analysis
- MoUs defining access to data

Implementation examples

- IceCube:regular data release since 2008. Specific sets for point- and diffuse searches
- Antares: specific point sources datasets
- HESS: released a test sample ~28 hours with a few sources in FITS format + IRF
- MAGIC: high-level published results and datasets in low-level format + IRF
- VERITAS: high-level data and sky maps in FITS file format
- HAWC: 1523-Day Survey catalog and specific analyses
- GW (LIGO VIRGO GEO600 KAGRA): all data available

Auger data sharing in the multi-messenger context

- MoU for cooperation with IceCube, Telescope Array, ANTARES & LIGO VIRGO collaborations for sharing of particular data sets
 - $\rightarrow\,$ shared publications and active working groups
- MoU with AMON consortium for real time streaming of sub-threshold events
 - $\rightarrow\,$ search for temporal and spatial correlations
- GNC Alarms are listened to and are routinely reported
 - $\rightarrow\,$ search for candidate neutrinos or photons
- Newborn consortiums as ACME (Astrophysics Centre for Multimessenger studies in Europe) funded up to 2028
 - $\rightarrow\,$ coordinating different institutions and human power
 - $\rightarrow\,$ exchange of information in quasi-real time
 - $\rightarrow\,$ develop joint analyses programs upon request



The Pierre Auger Observatory

\sim 400 members. 17 countries

Radio detector 153 Radio Antenna (AERA) Radio antenna array (153 antennas, 17 km²) LIDARs and laser facilities Sub-array of 750 m (63 stations, 23.4 km²) Underground muon detectors (24+) High elevation telescopes (3) 4 fluorescence detectors (24 telescopes up to 30°) entra ampus 1665 surface detectors: radio) water-Cherenkov tanks (grid of 1.5 km, 3000 km²)

Surface detector

array of 1660 Cherenkov stations on a 1.5 km hexagonal grid of 3000 km² Dense array SD750 of 24 km²

Fluorescence detector

4+1 buildings overlooking the array (24 + 3 HEAT telescopes)

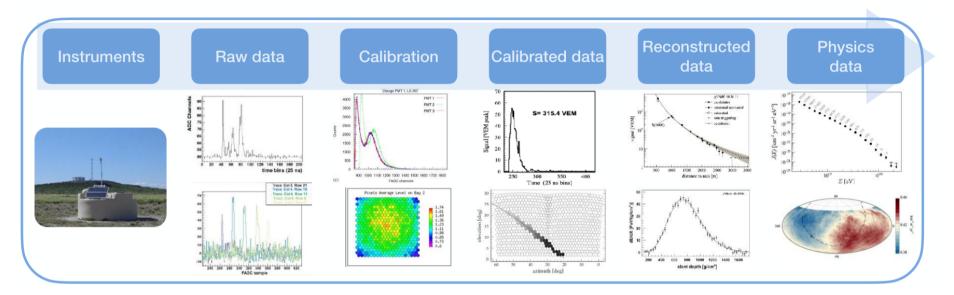
Muon Detectors

Buried scintillators (dense array)

Phase 1: 2004 - 2022

Phase 2 - the AugerPrime upgrade Data taking from 2023 to 2035.. Multiple detectors (scintillators - muon -

Data flow chart



 many different instruments: from main detectors to monitoring devices
 a large variety of data: from primary DAQ data to calibrated and reconstructed data suitable for physics analyses

 \rightarrow different competences needed for designing and implementing a data release policy!

The Data Release Task

Observatory Management

data management plan (production - preservation - release)

release policy (what - when - how much)

synergy with physics tasks coordinators

- identify suitable samples
- select and validate data
- reproduce analyses (following publication policy)
- regular maintenance and updates

organization:

- \rightarrow wikipage & mailing list
- \rightarrow monthly/biweekly online meetings

Open Data Policy: general thoughts

https://opendata.auger.org/AugerOpenDataPolicy.pdf

"The Pierre Auger Collaboration is committed to the public release of their data, at different levels of complexity, as well as of software tools developed for analysis, for the purpose of re-use by a wide community including professional scientists, educational and outreach initiatives, and citizen-scientists in the general public"

'as open as possible, as closed as necessary'

- increasing fraction of cosmic ray data collected by completed detectors
- **all** atmospheric data & space-weather data
- MC simulations and software tools
- $\rightarrow\,$ the policy is implemented through the definition of data levels
- \rightarrow the entire process is subject to approval by the Collaboration Board

Open Data Policy: current implementation

https://opendata.auger.org/AugerOpenDataPolicy.pdf

Data levels:

- **1.** Open access publications and additional numerical data \rightarrow at the moment of publication
- 2. Simplified data for education and outreach → 10% cosmic-ray data are released regularly in a simplified format. 100% of space-weather and atmospheric data
- 3. Reconstructed data / simulation + software & documentation \rightarrow 10% cosmic-ray data released (used for publications and in last ICRC)
- **4.** Close-to-raw data + software & documentation → public data releases comprising data used for publications and in last ICRC

Phase I data (Jan 2004 – Dec 2022)

- SD-1500 array and SD-750 array
- FD (hybrid) events used for calibration, spectrum & composition analyses
- Other-than-cosmic-ray-data: weather station data, scaler data, ELVES

From first steps to current status

2007: Public Browser

1% of Surface Detector data for educational purposes

Feb. 2021: Open Data Portal

10% cosmic ray data:

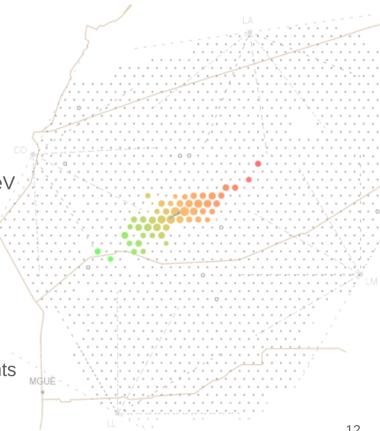
- ~ 6000 km² sr yr
- ~ 25000 events measured with the surface detector above 2.5 EeV & Hybrid events above 10^17.8 eV

CSV and JSON format including high-level & RAW info

100% atmospheric and space weather data

2021-2024:

- regular updates with extended data samples and code (inclined events and low energy extension) \rightarrow > 80000 events
- UHECR catalog published
- outreach section included



The Portal

• Datasets

Description of CR physics, detection principle and file content

• Visualization

Event selection and 3D browser

• Analysis

Python code for exploring data & understanding physics results

UHECR Catalog

Details of the 100 Highest energy events

• Outreach

Section dedicated to the general public

https://arxiv.org/abs/2309.16294

Pierre Auger Observatory Open Data

The Pierre Auger Open Data is the public release of 10% of the Pierre Auger Observatory cosmic-ray data published in recent scientific papers and at International conferences, following the Auger Collaboration Open Data Policy. The release also includes 100% of weather and space-weather data collected until 31 December 2020. This website hosts the datasets for download. Brief overviews of the <u>Pierre Auger Observatory</u> and of the <u>Auger Open Data</u> are set out below. 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.

Datasets Datasets the released datasets and their complementary data	O Visualize an online look at the released pseudo raw cosmic-ray data	Analyze example analysis codes in online python notebooks to run on the datasets	Catalog of the highest- energy cosmic rays	Dutreach a.page dedicated to the general public	
Open Data		Portal	Zenod	Zenodo DOI	

The Portal

The Pierre Auger Collaboration, Eur. Phys. J. C 85 (2025) 70

• Datasets

Description of CR physics, detection principle and file content

• Visualization

Event selection and 3D browser

• Analysis

Python code for exploring data & understanding physics results

• UHECR Catalog

Details of the 100 Highest energy events

• Outreach

Section dedicated to the general public

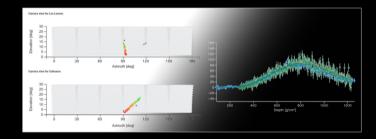
The European Physical Journal

volume 85 · number 1 · january · 2025

EPJC Recognized by European Physical Society

Particles and Fields

From "The Pierre Auger Observatory open data" by Pierre Auger Collaboration, Eur. Phys. J. C 85, 70 (2025).



Visualization of an exemplary event. 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.





Datasets

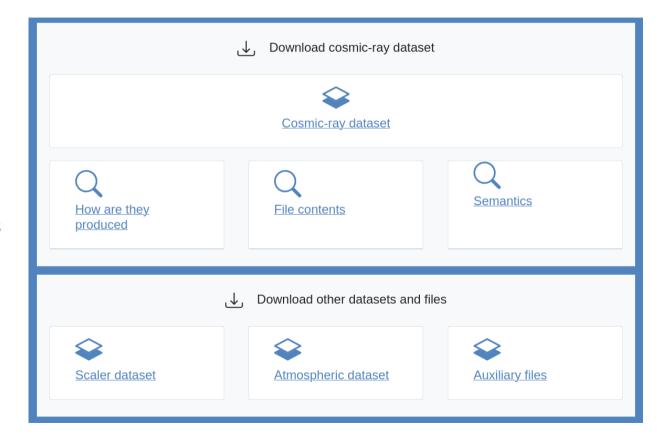
Data set description:

How data are **produced**

How data are **selected**

File **content** and semantics

Further references



Visualization

Event selection and browsing

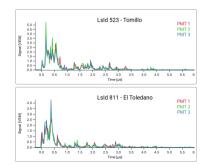
visualize exemplary events apply interactive selection

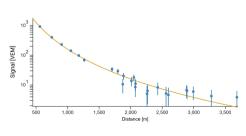
- Surface detector ADC traces _
- FD telescopes sky view _
- SD & FD reconstruction tabs _

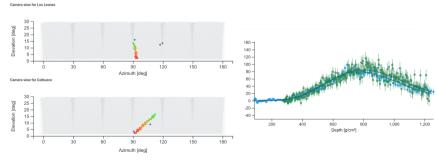
	Nb of stations		Energy [EeV]	Zenith Angle [deg]	Time [gps]	
Min.	1	\$	0	0	\$ 756950413	Ŷ
Max.		\$	1000	80	\$ 1261872018	~
Event type	SD Vertical	~	Select			

Visualize some example events >







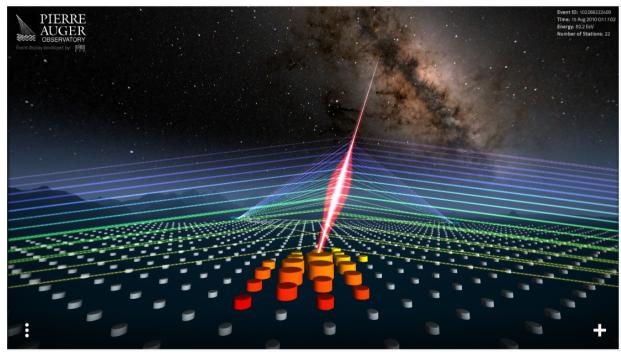


The most energetic hybrid event

Immersive and interactive view High-level info & raw data

PAO100815:

energy of 82 EeV zenith of 54 degrees triggered 22 SD stations fluorescence telescopes in 4 FDs



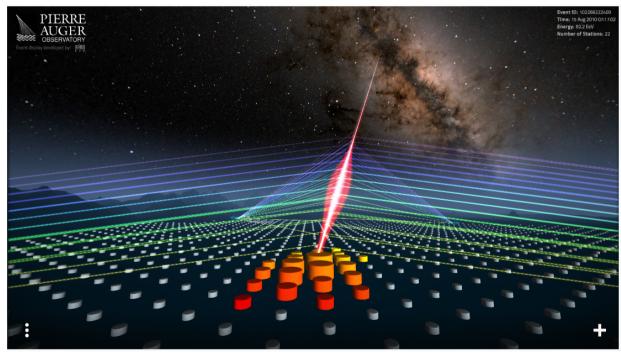
The most energetic hybrid event

Immersive and interactive view High-level info & raw data

PAO100815:

energy of 82 EeV zenith of 54 degrees triggered 22 SD stations fluorescence telescopes in 4 FDs

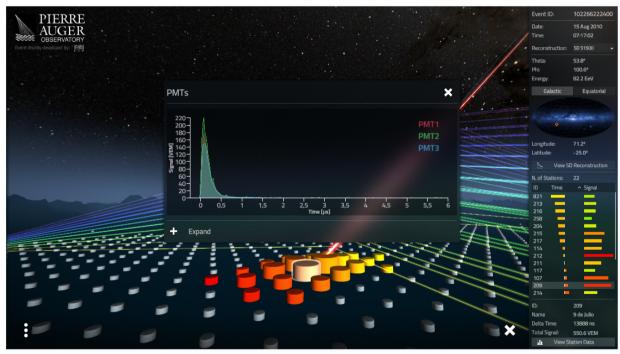
on my birthday!



The most energetic hybrid event

Immersive and interactive view High-level info & raw data

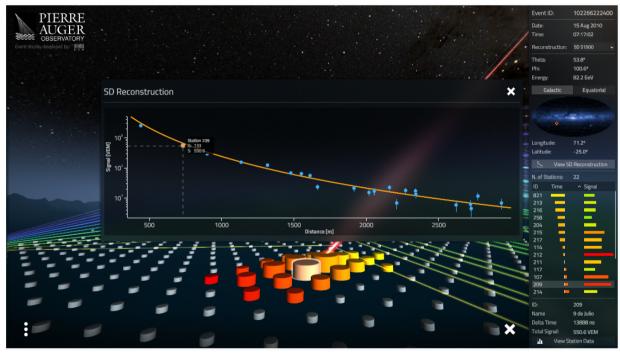
 \rightarrow SD station signals: calibrated traces for all PMTs in all the triggered stations



The most energetic hybrid event

Immersive and interactive view High-level info & raw data

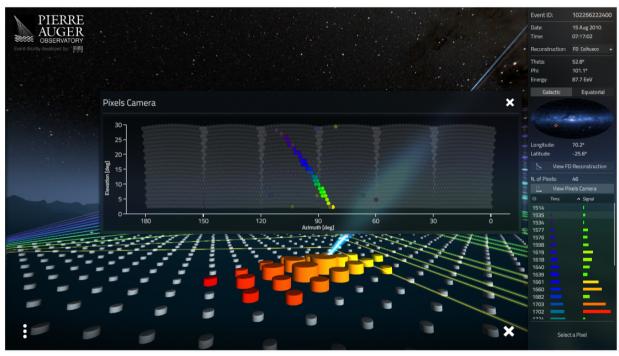
 \rightarrow SD reconstruction plot: LDF, signal falloff as a function of distance from shower core



The most energetic hybrid event

Immersive and interactive view High-level info & raw data

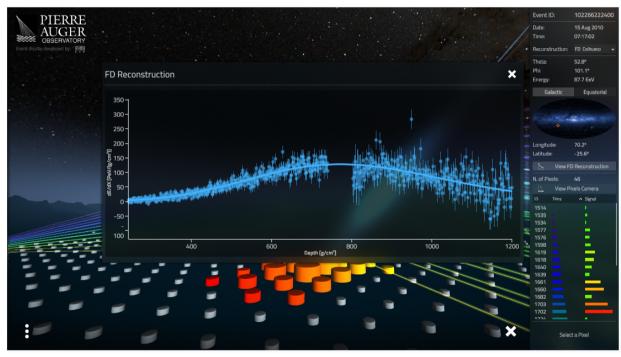
→ view from FD telescopes fluorescence cameras with hit pixels

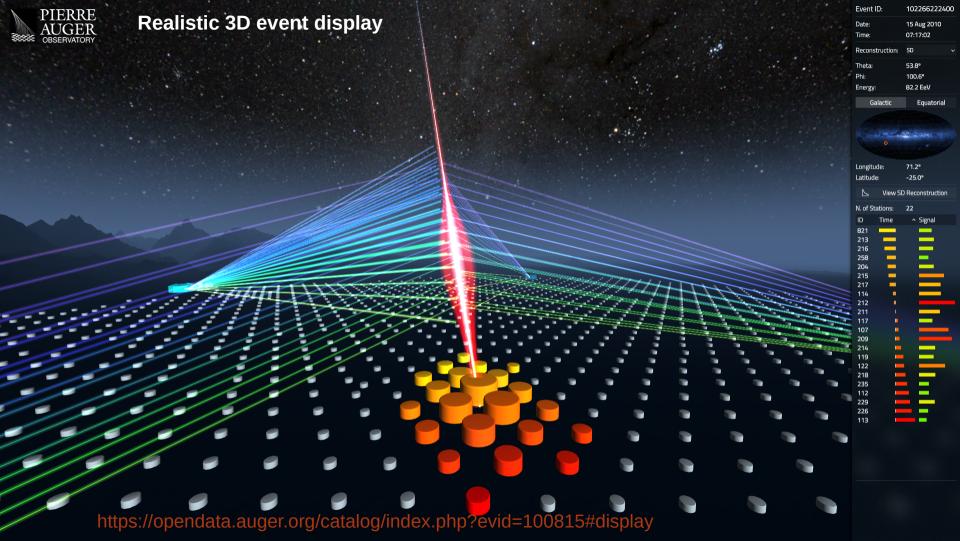


The most energetic hybrid event

Immersive and interactive view High-level info & raw data

→ FD reconstruction plot:
 Shower energy deposit
 as a function of atmospheric
 depth





Analysis tools

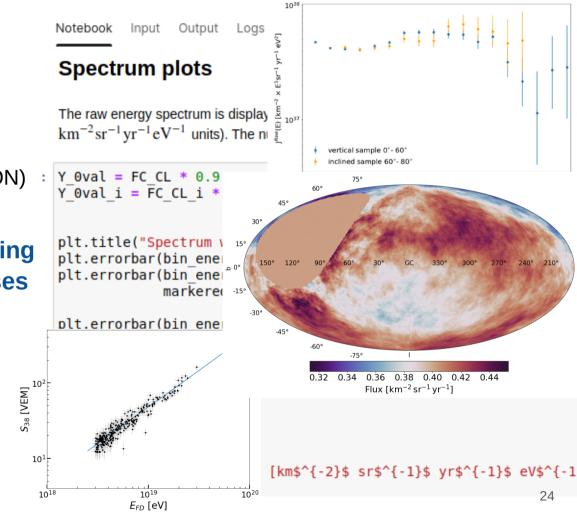
Tutorial codes (python)

- manipulate data (CSV and JSON)
- plot variables & histograms

Advanced codes for reproducing main physics results & analyses

- energy spectrum & calibration
- depth of shower maximum
- proton-air cross section
- UHECR sky
- weather corrections

→ run online or download



Analysis tools: energy spectrum

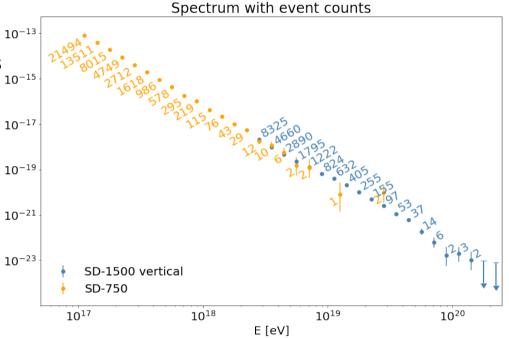
^{Raw}(E) [km⁻²

Phys. Rev. D 102, 062005 (2020) Eur. Phys. J. C 81, 966 (2021)

The raw energy spectrumCount the number of events in energy bins

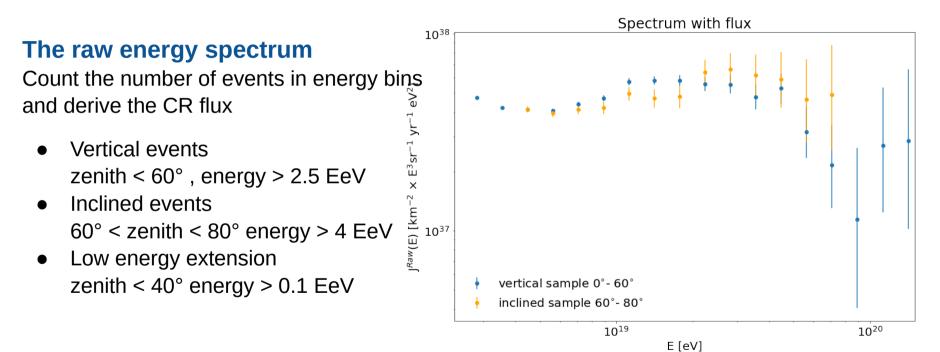
and derive the CR flux

- Vertical events zenith < 60°, energy > 2.5 EeV
- Inclined events zenith < 80° energy > 4 EeV
- Low energy extension zenith < 40° energy > 0.1 EeV



Analysis tools: energy spectrum

Phys. Rev. D 102, 062005 (2020) Eur. Phys. J. C 81, 966 (2021)



Analysis tools: energy calibration

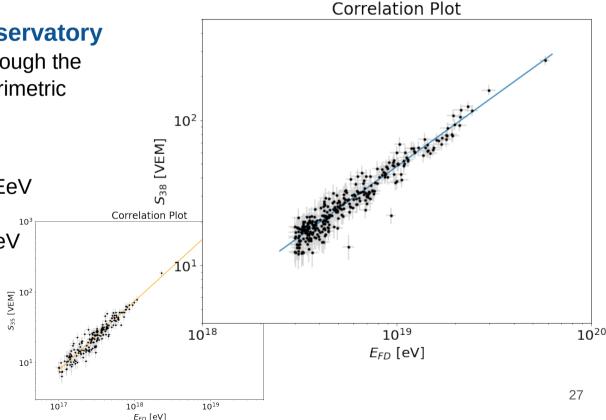
10

Phys. Rev. D 102, 062005 (2020) Eur. Phys. J. C 81, 966 (2021)

The energy scale of the Observatory

calibrate the energy estimator through the measurement of the shower calorimetric energy performed with the FD

- Golden hybrid events $zenith < 60^\circ$, energy > 2.5 EeV
- Low energy extension $zenith < 40^{\circ} energy > 0.1 EeV$



Analysis tools: depth of the shower maximum

Phys. Rev. D 90, 122005 (2014)

Xmax and elongation rate

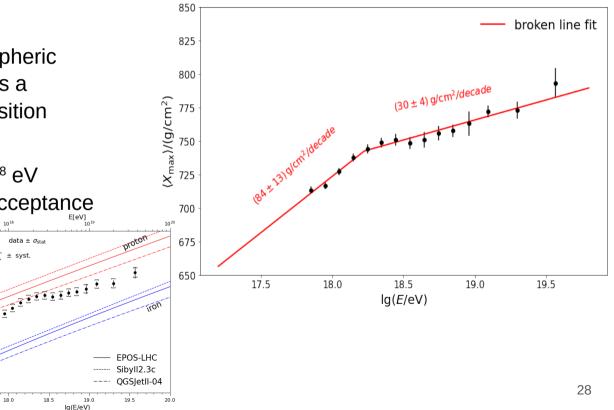
estimation of the average atmospheric depth of the shower maximum as a function of the energy \rightarrow composition

- Hybrid events zenith < 70°, energy > 10^{17.8} eV
- auxiliary files for detector acceptance and models predictions

800

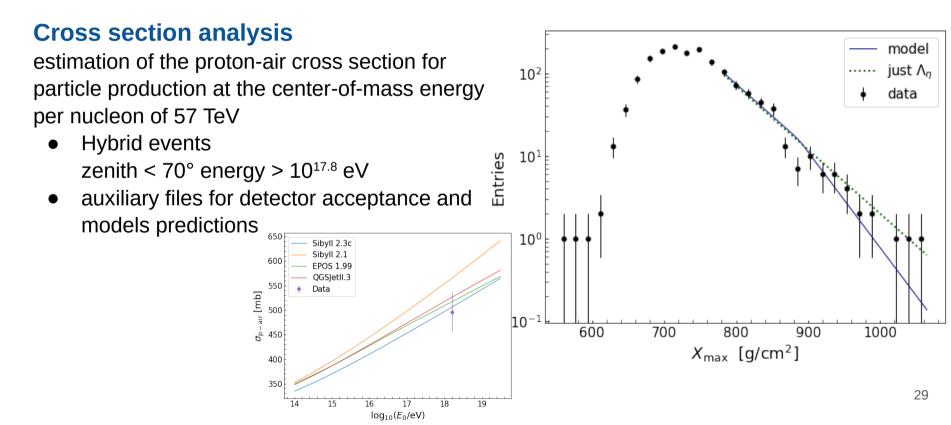
(X_{max}) [g/cm²]

600



Analysis tools: measurement of the cross section

Physical Review Letters 109 (2012) 062002



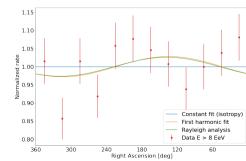
Analysis tools: the UHECR sky

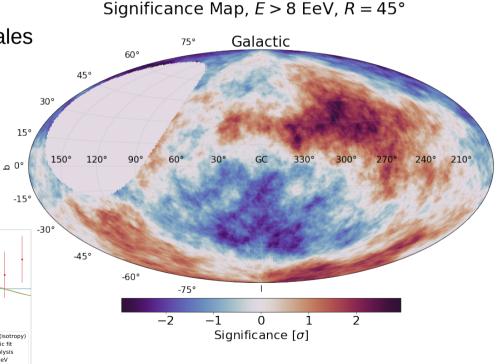
Science 357 (2017) 1266

Arrival direction analysis

search for anisotropies on large angular scales in the distribution of the arrival directions of cosmic rays

- Vertical events zenith < 60°, energy > 2.5 EeV
- Inclined events
 60° < zenith < 80° energy > 4 EeV





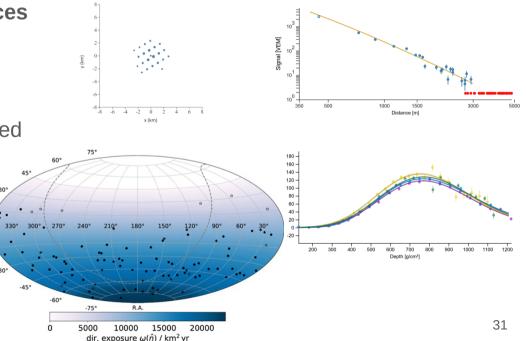
Catalog of the highest energy events

15 ق 0°

A. Abdul Halim et al 2023 ApJS 264 50

Highest energy CR events revealed

- High-level parameters and raw traces of the 100 highest-energy CRs recorded between 2004–2020
- Energies in the range 78–166 EeV.
- Further **9 very energetic events** used in the energy **calibration**
- sky plot of the arrival directions



Catalog of the highest energy events

A. Abdul Halim et al 2023 ApJS 264 50

32

Top 100 events: vertical

PAO191110

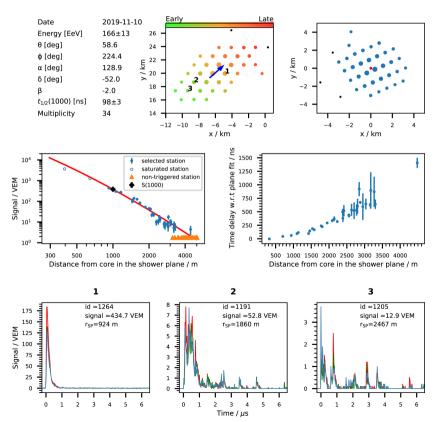
energy 166 +- 13 EeV

zenith 58°

34 triggered stations

footprint (13 × 6) km²

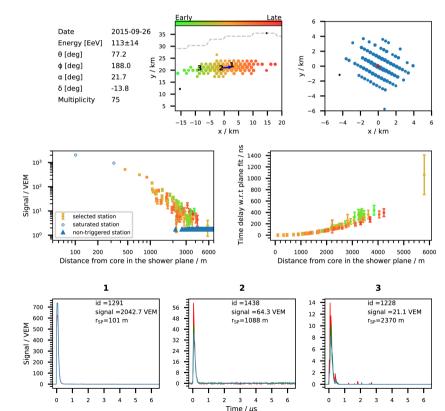
#1 - PAO191110



Catalog of the highest energy events

A. Abdul Halim et al 2023 ApJS 264 50

#17 - PAO150926



Top 100 events: inclined

PAO150926

energy 113 +- 14 EeV

zenith 77°

75 triggered stations

footprint (35 × 6) km²

Outreach section

Cosmic rays in a nutshell:

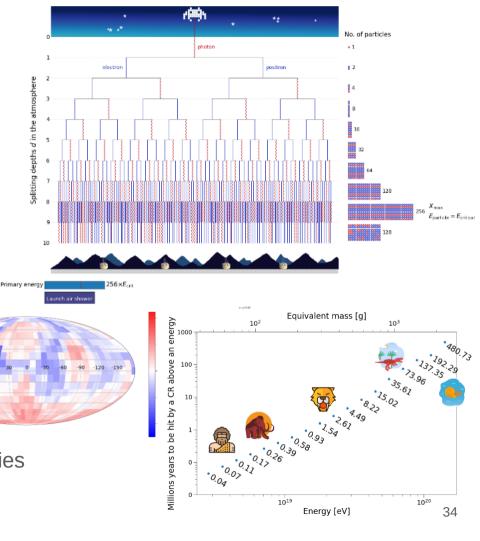
Simplified description of detection principles

Tutorials for learning data processing:

- Reading JSON & CSV files
- Basic plotting

Simplified phys. notebooks:

- plot weighted sky-maps
- understand shower development
- relate CR observables to daily-life quantities
- use weather and space-weather data



OpenData in international events

World-wide initiatives:

- IPPOG Masterclasses
- International Cosmic Day
- European Researchers' Night
- Hundreds of local initiatives
- > 60000 students 60 Countries



MASTERCLASSES

MASTERCLASSES

hands on particle physics

INTERNATIO

International Particle Physics Outreach Group

Discover Cosmic Rays

INTERNATIONAL COSMIC DAY

Cesmic particles, these unnoticed particles that surround us all the time are the focus of this day. Stadents, backers and scientists get together to talk and learn about Cesmic Rays and answer questions like:

> What are cosmic particles? Where do they come from? Now can they be measured? And what can we learn from them?

If you want to know more about the secrets they bring with and to be part of this collaboration, get here more information: inger Greit DEP (Namer Construction)

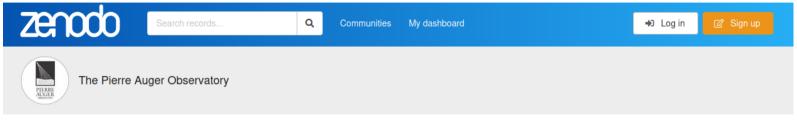
Mips://ed.desy.de Mips://www.laotbook.com/InternationalCosmixEng





Tracking use of OpenData: Zenodo

Zenodo DOI



🔒 Open

Dataset

Published 2024 | Version 3

.

Pierre Auger Observatory Open Data

The Pierre Auger Collaboration

The Pierre Auger Collaboration is releasing 10% of the data recorded since 2004 using the world's largest cosmic ray detector, the Pierre Auger Observatory, located in Argentina, in the Province of Mendoza. The release also includes 100% of weather and space-weather data collected until 31 December 2020. These data are being made available publicly with the expectation that they will be used by a wide and diverse community including professional and citizen-scientists and for educational and outreach initiatives.

Operation of the Pierre Auger Observatory, by a Collaboration of about 400 scientists from over 90 institutions in 18 countries across the world, has enabled the properties of the highest-energy cosmic rays to be determined with unprecedented precision. These cosmic rays are predominantly the nuclei of the common elements and reach the Earth from astrophysical sources. The data from the Observatory have been used to show that the highest-energy particles have an extra-galactic origin.

Cosmic rays are observed indirectly, through extensive air-showers of secondary particles produced by the interaction of the incoming cosmic ray with the atmosphere. The Surface Detector of the Observatory covers 3000 km² and comprises an array of ~1600 particle detectors, separated by 1500 m. The low energy extension features an array of 71 stations spread apart by 750 m and covering about 27 km². The area is overlooked by a set of telescopes that compose the Fluorescence Detector which is sensitive to the auroral-like light emitted as the air-shower develops, while the Surface Detector is sensitive to muons, electrons and photons that reach the ground



Versions	
Version 3 10.5281/zenodo.10488964	2024
Version 2.0 10.5281/zenodo.6867688	Dec 22, 2022
Version 1.1.0 10.5281/zenodo.5588460	Oct 26, 2021
Version 1.0.0 10.5281/zenodo.4487613	Feb 15, 2021

Tracking use of OpenData: Matomo

30

15

0 Thu, lul 1

Sun, Dec 19

Wed, Jun 8

📣 matomo

server tracking since 7/2021

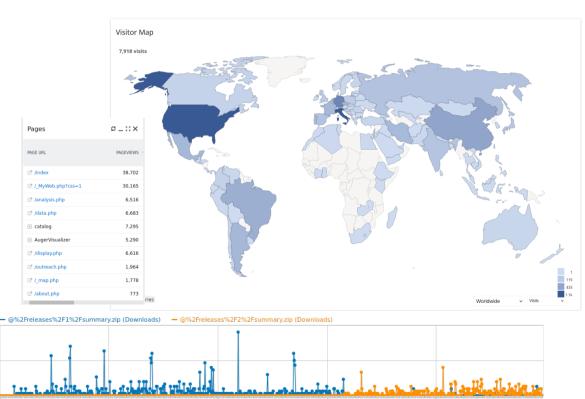
- world-wide visit map
- pages visits monitoring
- downloads tracking

metrics

 ~ 70000 visits

(8000 visits av.time >1 min

> 4300 data downloads



Sat, Nov 26

Tue, May 16

Papers citing OpenData

ArXiv and PRD

- Composition studies
- Hadronic models
- Effect of IMF

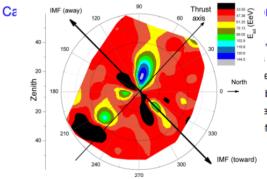
Learning the composition of ultrahigh energy cosmic rays

Blaž Bortolato, Jernej F. Kamenik, and Michele Tammaro Phys. Rev. D **108**, 022004 – Published 13 July 2023



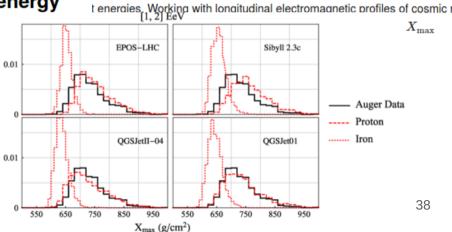
[Submitted on 17 Jan 2023 (v1), last revised 20 Jan 2023 (this version, v2)]

Alignment of air showers produced by ultra-high energy cosmic rays at the Pierre Auger Observatory



muceno

, azimuth) distribution of extensive air s at the Pierre Auger Observatory (PAO) eld (IMF), with a thrust value $Tp \geq 0.6 \frac{1}{et}$ behavior strongly suggests an effect of eral scattering. We discuss the weakeni f observational time) when the IMF bec



erence on the Pierre Auger Open Data to discern the mass composition

Increase of the public CR data

Next release planned for late 2025

New policy approved by the Collaboration Board

 $\rightarrow\,$ increase cosmic ray data fraction to 30%

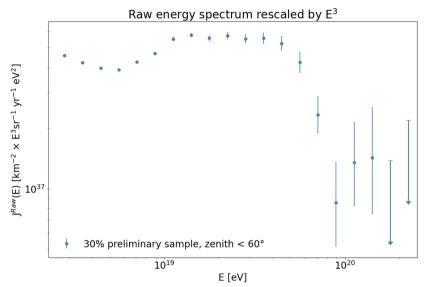
first SD1500 vertical events > 2.5 EeV

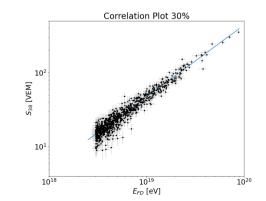
+ hybrid events used for energy calibration

→ unprecedented exposure ~ 24000 km² sr yr!

To be presented at the ICRC 2025







Summary and Outlook

data sharing key point for present experiments and observatories

Funding agencies – many projects and initiatives worldwide – human power required!

source for multi-messenger campaigns

Synergy among experiments - shared analysis programs

large impact in dissemination and outreach

Many initiatives dedicated to general public

Pierre Auger Open data Portal

Setup in 2021, since then regularly updated and extended released CR fraction increases to 30% of Phase I data (~24000 km² sr y) Phase II: new data and new detectors

Thanks!