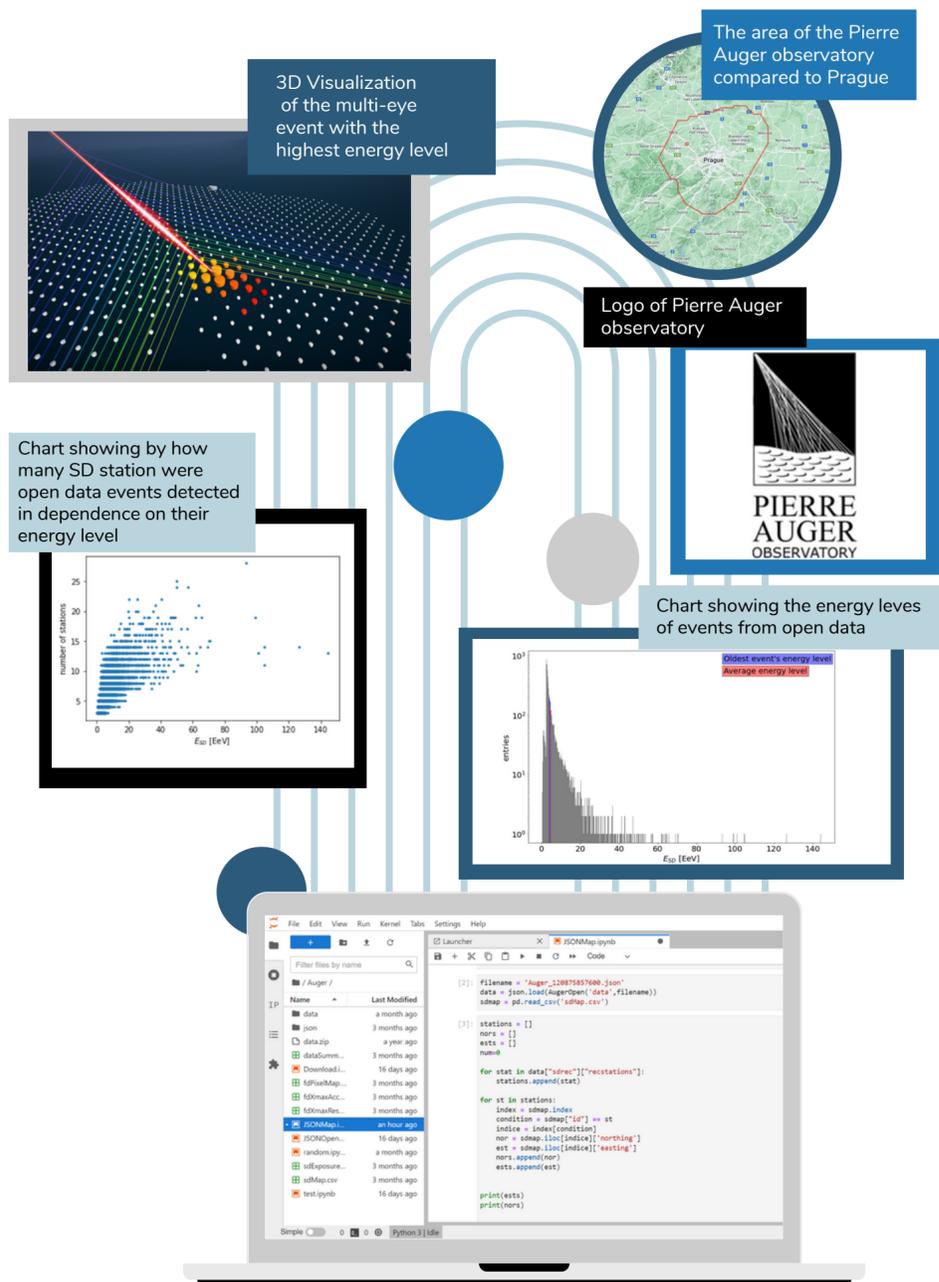


# ANALYSIS OF PIERRE AUGER OBSERVATORY OPEN DATA USING EGI JUPYTER NOTEBOOKS

Secondary school students during their internship in the Czech Academy of Sciences compare the Kaggle platform with usage of EGI Jupyter notebooks while analyzing open data published by the Pierre Auger Observatory.



## EGI JUPYTER NOTEBOOKS

EGI Jupyter notebooks is an on-line environment based on EGI cloud service, which can be used as a tool for data analysis. To be able to connect to and work in the default version of this environment it is necessary to have an online profile which can be created for example through Google account for free. The default version is limited to 2 CPU cores, 4 GB RAM and 20 GB of persistent storage per user. The notebooks are ready to use in under a minute as well. We had a problem importing the Pierre Auger open data via the import button, therefore we had to use a terminal to download data.

## CONCLUSION

EGI Jupyter notebooks are a good and suitable alternative to the Kaggle platform. Their web-based environment is easy to navigate and work in. Apart from the fact that the Pierre Auger Observatory data must first be manually downloaded in order to be analyzed, we have not come across any downsides of this tool. For our analyses and general use, all the aspects of EGI Jupyter notebooks were satisfactory.

## AUTHORS



Václav  
Zajac  
SPSŠT  
Panská



Filip  
Neubauer  
Akademické  
gymnázium Štěpánská



Jana  
Maršalková  
Gymnázium  
Jana Nerudy



Jiří  
Chudoba  
supervisor

## PIERRE AUGER OBSERVATORY

The Pierre Auger Observatory (PAO), located in pampa in Argentina province Mendoza, measures showers of particles induced by ultra-high-energy particles originated in space. The uniqueness of the observatory lies not only in its large detection area of 3000 km<sup>2</sup> but also in its usage of the combination of several detecting methods. The observatory uses simultaneously two types of modern detectors to collect information about cosmic rays: the surface detectors (SD) and the fluorescence detectors (FD). Even with the utilization of this "hybrid" detector, studying the highest energy cosmic rays is very challenging as they are extremely rare. Due to this fact the origin of these high-energy particles remains unknown.

## OPEN DATA

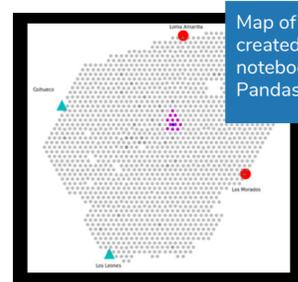
The Pierre Auger Observatory 2021 open data contain information about 22731 showers recorded with surface detectors together with data from 3156 hybrid events which were recorded with both surface and fluorescence detectors. In this dataset we can find 10 % of all the events recorded by the Pierre Auger Observatory from 2005 to 2020. The data are available in pseudo-raw JSON files which include all the information available about each shower event and in CSV dataset containing its summary.

## OPEN SCIENCE

Open science is a project created by the Czech Academy of Sciences that allows secondary school students to participate in scientific research through internships. Last year Open science received 1584 applications from 650 students for 125 internships.

## KAGGLE

Kaggle is an online community platform for data science and machine learning on which every registered user can publish and find numerous datasets including the PAO open data. It is possible to download the data on a local computer or work with them directly on the website as Kaggle provides Jupyter notebooks. Available CPU offers 4 cores, 16 GB RAM and 20 GB of auto-saved disk space. The usage limit of the CPU is 12 hours per session. The notebooks start and are ready to use in under a minute and they automatically stop after 40 minutes of no activity. Working with small datasets we did not encounter any problems while using Kaggle notebooks.



Map of multi-eye event created on EGI Jupyter notebook using Python Pandas



Photo of the surface detector

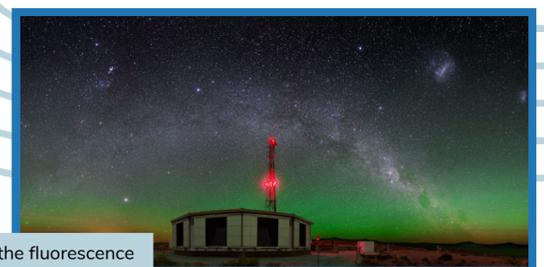


Photo of the fluorescence detector in night

sources: [www.auger.org](http://www.auger.org), [www.opendata.auger.org](http://www.opendata.auger.org)