Data open-access policy of the Pierre Auger Observatory

The Pierre Auger Observatory (Auger), located in Argentina, in the Province of Mendoza, is the world's largest cosmic ray observatory. The Collaboration comprises about 400 scientific authors from more than 90 institutions, representing 18 countries from all over the world. The Collaboration is led by two spokespersons, elected by the Collaboration Board, which is the policy- and decision-making body. The Project Manager, assisted by a Deputy Manager, is responsible for the construction and operation of the Observatory and reports to the Spokespersons.

This policy describes the Auger principles of data open-access. The data from Auger are the result of vast and long-term moral, human and financial investment by the international community. There is a unique scientific opportunity in re-using these data, at different levels and at different points in time. 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. The Collaboration upholds the principle that open access to the data and the software will, in the long term, allow the maximum realization of their scientific potential.

Auger data come from a variety of instruments and take many forms, starting from either raw experimental or simulated data through reconstructed data and datasets of higher level generated by analysis workflows, all the way up to data presented in scientific publications. In the field of high energy physics, four levels of complexity of data have been identified[1]. The Auger Collaboration has adopted this classification and adapted it to its context, for its open-access policy:

Level 1: Open access publications and additional numerical data.

Auger Level 1 Policy: Scientific results are made publicly accessible by publishing either in open-access journals or in the arXiv or in both. The Collaboration strives to provide additional numerical information to facilitate immediate re-use, this information being provided either as tables in the publications themselves or as downloadable files at the official Auger website[2].

Level 2: Simplified data for education and outreach.

Auger level 2 policy: Reconstructed cosmic-ray events, and close-to-raw data for each of them, in a simplified format are made publicly available for education and outreach purposes[3]. Space-weather and atmosphere-monitoring data are also released.

Level 3: Reconstructed data and simulation together with example software (and documentation) to analyse them.

Auger level 3 policy: Reconstructed cosmic-ray events and simulations, selected with the best available knowledge of the detector performance and conditions of the data-taking, are made publicly available. An event-display, as well as example codes to process and analyse

1 C. Diaconu et al., "Data Preservation in High-Energy Physics" http://arxiv.org/abs/0912.0255

2 https://www.auger.org/index.php/science

3 https://labdpr.cab.cnea.gov.ar/ED/index.php

level 3 data, are also provided. The codes are derived from those used by the Collaboration for published analyses.

Level 4: Close-to-raw data with example software (and documentation) to access and understand them.

Auger level 4 policy: Close-to-raw data associated with the cosmic-ray events released at level 3 are made publicly available. An event-display, as well as codes to read level 4 data, are also provided.

The Auger Collaboration will provide open access to different kinds and fractions of data, at different points in time and with appropriate delays, which will allow Auger collaborators to fully exploit the scientific potential of the data before Open Access takes place.

The current Auger policy, as approved by the Auger Collaboration Board, foresees that:

At level 1, the additional numerical data are made available at the moment of the publication.

At level 2, 10% of the cosmic-ray data are released regularly in a simplified format. 100% of the space-weather and atmospheric data are also made public.

At level 3 and 4, public data releases, accompanied by open source software and suitable documentation, take place regularly. The fraction of the data which Auger will first make available is 10% of data samples used for physics results as presented at the preceding International Cosmic Ray Conference.

The release of level 2 cosmic-ray data (1% of them) from one of the instruments of the Observatory, the Surface Detector 1500 m array (SD-1500), has been taking place regularly since 2007, as a stress-test exercise of the Open Access chain. The release also includes 100% of the space-weather data. The fraction of the level 2 cosmic-ray data was increased to 10% in 2019, after which the Collaboration Board endorsed, in 2020, the release of the same fraction of level 3 and 4 data from the SD-1500 and from another instrument, the Fluorescence Detector (FD). The first release of these level 3 and 4 data, accompanied by appropriate simulated data, is taking place in 2021.

Approved by the CB on Jan 28 2021