

The KASCADE Cosmic Ray Data Centre (KCDC)

S. Schöo^{*1}, J. Blümer^{1,2}, B. Fuchs², A. Haungs¹, D. Kang², D. Wochele¹, J. Wochele¹, W.D. Apel¹, J.C. Arteaga-Velázquez³, K. Bekk¹, M. Bertaina⁴, H. Bozdog¹, I.M. Brancus⁵, E. Cantoni^{4,6}, A. Chiavassa⁴, F. Cossavella², K. Daumiller¹, **V. de Souza⁷**, F. Di Pierro⁴, P. Doll¹, R. Engel¹, D. Fuhrmann⁸, A. Gherghel-Lascu⁵, H.J. Gils¹, R. Glasstetter⁸, C. Grupen⁹, D. Heck¹, J.R. Hörandel¹⁰, D. Huber², T. Huege¹, K.-H. Kampert⁸, H.O. Klages¹, K. Link², P. Łuczak¹¹, H.J. Mathes¹, H.J. Mayer¹, J. Milke¹, B. Mitrica⁵, C. Morello⁶, J. Oehlschläger¹, S. Ostapchenko¹², N. Palmieri², T. Pierog¹, H. Rebel¹, M. Roth¹, H. Schieler¹, F.G. Schröder¹, O. Sima¹³, G. Toma⁵, G.C. Trinchero⁶, H. Ulrich¹, A. Weindl¹, J. Zabierowski¹¹ -

KASCADE-Grande Collaboration

¹ Institut für Kernphysik, KIT - Karlsruhe Institute of Technology, Germany

² Institut für Experimentelle Kernphysik, KIT - Karlsruhe Institute of Technology, Germany

³ Universidad Michoacana, Inst. Física y Matemáticas, Morelia, Mexico

⁴ Dipartimento di Fisica, Università degli Studi di Torino, Italy

⁵ Horia Hulubei National Institute of Physics and Nuclear Engineering, Bucharest, Romania

⁶ Osservatorio Astrofisico di Torino, INAF Torino, Italy

⁷ Universidade São Paulo, Instituto de Física de São Carlos, Brasil

⁸ Fachbereich Physik, Universität Wuppertal, Germany

⁹ Department of Physics, Siegen University, Germany

¹⁰ Dept. of Astrophysics, Radboud University Nijmegen, The Netherlands

¹¹ National Centre for Nuclear Research, Department of Astrophysics, Lodz, Poland

¹² Frankfurt Institute for Advanced Studies (FIAS), Frankfurt am Main, Germany

¹³ Department of Physics, University of Bucharest, Bucharest, Romania

E-mail: sven.schoo@kit.edu

KCDC is a platform on the web for distributing air-shower events measured by the KASCADE and KASCADE-Grande experiments. In addition to the data, an extensive documentation on the experiments and the published observables is provided to ease the use of the data in analyses performed by the cosmic ray community. Since the experiments have been funded by tax-payers, access is also open to the broad public. Therefore, additional information and example analyses are needed and provided to enable students and pupils to get in touch with astroparticle physics. In its current release, KCDC provides access to 160 million events measured by the KASCADE array and central detector. The data set consists of 18 observables including both, reconstructed shower properties, like the number of charged particles at ground, and meta data, like when the event has been recorded. In the first year of operation nearly 100 users from more than 30 countries distributed over 5 continents logged on to the KCDC web-platform. This contribution will cover the current state of the project and the future plans for extending the published data set and the software implementation of KCDC.

*The 34th International Cosmic Ray Conference,
30 July- 6 August, 2015
The Hague, The Netherlands*

^{*}Speaker.

1. Introduction

Having measured for more than 20 years, KASCADE [1] has recorded more than 1.7 billion air-showers induced by high energetic particles ($10^{14} - 10^{17}$ eV) in the atmosphere. Some 425.000.000 survive all quality cuts. Around 160.000.000 events are already publicly available.

Next to the publication of the recorded data, the short term aim of the project KCDC – KASCADE Cosmic-ray Data Centre [2] – is to provide the public with a detailed documentation of the experiments components ($200 \times 200 \text{ m}^2$ area containing shielded and non-shielded detectors, large-size hadron calorimeter, and additional muon tracking devices), of the reconstruction procedures, and of the observables that are being published.



Figure 1: Logo of KCDC. The link to KCDC is: <https://kcdc.ikp.kit.edu>.

In accordance with the ‘Berlin Declaration on Open Data and Open Access’ [6] which explicitly requests the use of web technologies and free, unlimited access for everyone, KCDC is being implemented as a web portal.

In the future, KCDC will be extended not only in terms of the number of events and observables, but also in terms of experiments. As its extension, the KASCADE-Grande array coexisted for a long time next to the KASCADE array. The events measured by KASCADE-Grande [3] on an area of $700 \times 700 \text{ m}^2$ within an energy-range of $10^{16} - 10^{18}$ eV will be added to KCDC as well as data recorded by the radio antenna field LOPES [4] and the microwave experiment CROME [5].

2. The Web Portal KCDC

The web portal (entrance page see Figure 3) is the access-point for the user to the documentation and to the data of the KASCADE experiment.

We declared both, the scientific and the non-scientific audience as focus of possible users. To ease the access and usage especially for non-physicists, we included some tutorials on working with the platform and the data (see section 5).

It is foreseen that the software behind the data centre including the web portal is also made available as open source at a later stage. Therefore, we were anxious that KCDC, in addition to fulfilling the needs to publish the data, provides a modern software solution that is understandable

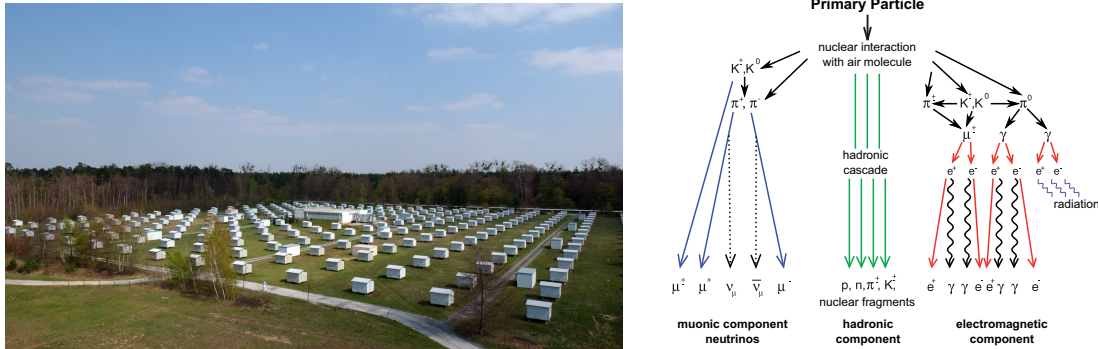


Figure 2: Photograph of the KASCADE experiment (left panel); Schematic view of an extensive air shower (EAS), where KASCADE is measuring the hadronic, muonic, and electron components (right panel).

for a general audience. This applies to the handling of the software by the end user as well as to the source code that could be the basis for potential developers.

KCDC is implemented as a plugin system, therefore, it is easy to add, remove, or modify content or features in order to adapt the framework to the needs specific to other experiments or other kinds of data.

3. The Data Shop

If registered via the ‘user page’, the user is able to enter the data shop. The registration is necessary in order to ensure that the ‘End User License Agreement’ is read, i.e. the legal aspects of



Figure 3: Title page of the website of KCDC, the KASCADE Cosmic-ray Data centre.

public data are understood (see also next chapter).

Figure 4 shows a screen shot of the current data shop with the list of parameters available per event. The looks will change in future releases once we added additional components of the experiment, however, the general idea will stay the same.

First the user has the chance to limit the number of observables that are being returned. If the air-pressure is not needed, it does not have to be shipped. However, there are two quantities that are always selected, the identifiers of the individual event, i.e. the number of the run and the number of the event of that run. For each quantity, a short description is shown on the right hand side together with a link to the user manual, which has a more detailed description of the observable.

Once the user has selected the observables, cuts may be applied. This is useful if one is only interested, for example, in inclined showers or in high energy events, which will drastically reduce the number of events and, therefore, the file size and the time needed to select events later during the analysis.

After a small summary of the selected observables and applied cuts, the user may submit the request. It is then processed, the current status being available on the user page. Once the data is ready for download, the user will get a notification email (if enabled) and the data is kept available via ftp for a small time period (currently 2 weeks).

The data is currently shipped as ASCII-files, which are easy to use without the necessity to use specific libraries. Since most physicists are likely to prefer either the root-format [7] or hdf5-format [8], these formats will be supported in the next release.

The data can be used for any analysis, presumed that the user accepts the ‘limited use licence’ (following text is taken from the KCDC-EULA, see also next chapter):

Subject to your agreement and continuing compliance with the KCDC Terms, KIT hereby grants to you a limited, personal, nonexclusive, non-transferable, non-assignable and fully revocable license to – (a) use the webportal and (b) download and use the scientific data of the KCDC in compliance with good scientific practice – provided through the webportal or related online services for your non-commercial scientific purposes only. Commercial purposes are defined as projects for your own or third parties for which you are paid or granted values in lieu of cash for the use of the data.

There is no restriction on the kind of analysis with the provided data nor the publication of the results. However, the KCDC team would acknowledge notification on a use exceeding private education, as well as bug reports or suggestions for improvements. This can be done directly via the web portal and/or per Email to ikp-kcdc@lists.kit.edu.

4. Legal Aspects of KCDC

Opposite to software open source publications, there is no standard procedure yet available for open data publication. In cooperation with KIT and its law department we developed an own license based on the EULA (end user license agreement) model [9], adapted from that one often used for software.

We had to consider a twofold issue as the license is needed for the web portal and the data. The KCDC approach is based on the EULA model, because it is flexible and adaptable to our needs,

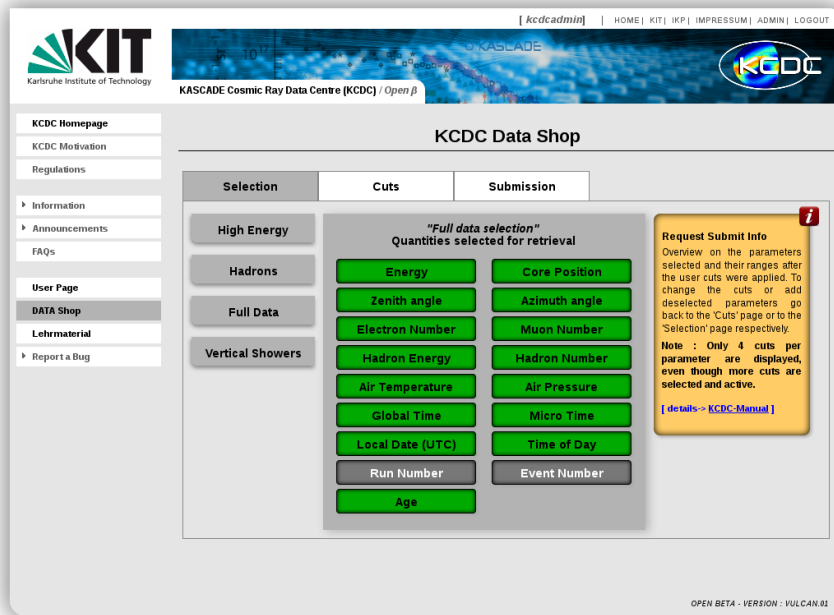


Figure 4: Screen shot of the ‘data shop’ of KCDC.

it includes the idea of requiring a good scientific practice, and it can be signed during registration and can be shipped with each data package.

In our custom-made adaption of the KCDC EULA we followed some key points from industry, like

1. no warranty for damage by owner of web portal or data;
2. no guarantee for availability or uptime of the server;
3. in case of disputes with local laws the EULA intention is conserved;
4. changes are possible at any time;
5. the termination of EULA is at our digression, only,

as well as obvious requirements from the open data idea, like

1. free access to the data and the web portal;
2. good scientific practice for the work with the data;
3. commercial usage of the data is not prohibited¹;

¹Please note, that the apparent contradiction to the statement above concerning the non-commercial use of the data, is solved by following statement in the EULA: ‘As an alternative to this EULA, KIT offers a license to use the DATA commercially as well, on the basis of a commercial license agreement. If YOU are interested in such a license please contact KIT, Institute for Nuclear Physics (IKP), the KCDC Group and the contact person provided on the WEBPORTAL’.

4. the citation of collaboration, KIT, and the web portal is mandatory;
5. free redistribution of data 'as is'.

5. Tutorials

The goal of having detailed tutorials, i.e. an 'education portal', is to provide the data also to a general public in the sense of a visible outreach of astroparticle physics. The first tutorials are already available, but presently in German, only and limited to two examples. English and also other languages can be added without any problem and it will be done in the near future.

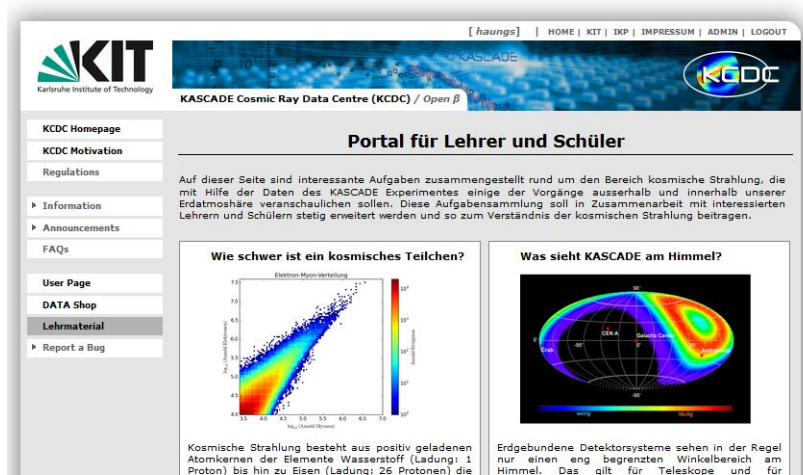


Figure 5: Impression of the tutorial section of KCDC.

The target group for the tutorials are teachers and pupils in high schools. This means a tutorial has to provide

1. a basic knowledge on KASCADE, astrophysics and related topics;
2. the required software and KCDC data (preferably as a pre-selection);
3. a step by step explanation of a simple data analysis;
4. a modern programming language code example;
5. the interpretation and discussion of the outcome.

This part we develop together with local teachers and pupils.

6. Future Steps

Several tasks and ideas are on the working list for further development of KCDC (some of them are agreed on, others are under discussion):

- Extending the educational portal with more examples and in various languages

- Improving the data selection process to be more comfortable and faster
- Improving the capabilities of the data-base and corresponding software
- Adding more data of KASCADE per individual event, such as particle densities per station
- Adding more events from KASCADE
- Adding simulated data
- Adding data of other KASCADE-Grande components like the Grande array or LOPES
- Adding data of other cosmic ray experiments
- Publication of the software behind KCDC
- Inclusion of KCDC in long-term data archive networks, e.g. Re3data [10],
-

Notice

Please note that since this article has been written, KCDC was further developed and improved. The description here is based on the *Open Beta Version: VULCAN.01*. Please visit the KCDC page and click the button ‘Developer News’ for more information.

References

- [1] Antoni T et al., *The Cosmic-Ray Experiment KASCADE*, *NIM A* **513** (2003) 490
- [2] see <https://kcdc.ikp.kit.edu>, Mai 2015
- [3] Apel W D et al., *The KASCADE-Grande experiment*, *NIM A* **620** (2010) 202
- [4] Falcke H et al., *Detection and imaging of atmospheric radio flashes from cosmic ray air showers*, *Nature* **435** (2005) 313
- [5] Smida R et al., *Observation of microwave emission from extensive air showers with CROME*, *EPJ Web Conf.* **53** (2013) 08010
- [6] see http://en.wikipedia.org/wiki/Berlin_Declaration_on_Open_Access_to_Knowledge_in_the_Sciences_and_Humanities, Mai 2015
- [7] Rene Brun and Fons Rademakers, *ROOT - An Object Oriented Data Analysis Framework*, 1997 *NIM A* **389** 81-86
- [8] The HDF Group. Hierarchical Data Format, version 5, 1997-2015.
<http://www.hdfgroup.org/HDF5/>
- [9] see http://en.wikipedia.org/wiki/End-user_license_agreement, Mai 2015
- [10] see <http://www.re3data.org/>, Mai 2015