PROCEEDINGS OF SCIENCE

KCDC – the KASCADE Cosmic-ray Data Centre

PoS(ICRC2023)1614

Andreas Haungs,[a,*] **Donghwa Kang,**[a] **Katrin Link,**[a] **Frank Polgart,**[a] **Sven Schoo,**[b] **Victoria Tokareva,**[a] **Doris Wochele**[a] **and Jürgen Wochele**[a]

[a] *Karlsruhe Institute of Technology, Institute for Astroparticle Physics, 76021 Karlsruhe, Germany*
[b] *Schoo Software Engineering, 76474 Au am Rhein, Germany*

E-mail: andreas.haungs@kit.edu

The 'KASCADE Cosmic-ray Data Centre', KCDC (https://kcdc.iap.kit.edu), is a platform where in first order the scientific data from the completed air-shower experiment KASCADE-Grande is made available for the astroparticle community as well as for the interested public. In addition, KCDC offers users some interesting features and services that have been successively added in recent releases. In the past two years, we have improved performance in particular with a new mongoDB and the update of all relevant software, necessary to make the KCDC data shop faster and more stable. In addition to corresponding simulations, this also concerns data from other experiments, as well as access to Jupyter notebooks. With the latest release end of 2022, we added a plugin which allows event displays for the array data of KASCADE-Grande measured air-showers as well as C++ programs to enable users to generate their own event displays. In this update we present a brief history of KCDC, the main features of the recent releases as well as a short outlook to future development plans.

38th International Cosmic Ray Conference (ICRC2023)
26 July - 3 August, 2023
Nagoya, Japan

[*] Speaker

1. Introduction

With KCDC [1] (figure 1), we provide the public the edited data, i.e. the reconstructed parameters of the primary cosmic rays measured by the detection of extensive air showers (EAS) with the KASCADE-Grande experiment – via a customized web page. The aim of this particular project was the installation and establishment of a public data centre for high-energy astroparticle physics. In the research field of astroparticle physics, such a data release was a novelty, whereas the data publication in astronomy has been established for a long time. KCDC provided the first conceptional design how the data of air-shower measurements can be treated and processed so that they are reasonably usable also outside the community of experts in the research field by the interested public [2]. The original idea and the realisation concept followed three requirements:

Figure 1: Logo of KCDC. The link to KCDC is `https://kcdc.iap.kit.edu`.

- When KASCADE-Grande was completed and dismantled, we wanted to make a secure repository of all scientific data, including the simulations and the software tools used, accessible to the members of the scientific collaboration and the scientists of the experiment.
- At the same time, we wanted to allow external experts access to the data to enable further use of the valuable data with possibly new insights and outcomes.
- Last, but not least, the platform should make open data available to the public, for open science, education, training and outreach.

In all our requirements we have followed the principles of FAIR Data Management [3].

Since we published the first KASCADE data in the KCDC DataShop in November 2013, it has been constantly expanded to include more data and other DataShops like KASCADE-Grande and Maket-ANI [4]. In the 10 years of KCDC we constantly published new major releases, where the news are documented at the web-based platform. In short:

- With the first release **WOLF359** in 2013, we published 158 Mio events with seven reconstructed parameters, based on data analyses of the original KASCADE detector array only.
- With the release **VULCAN** in 2014, we switched from SQL to NoSQL data base (MongoDB) and added data from the KASCADE Hadron Calorimeter.
- With **MERIDIAN**, released in 2015, the detector data from each of the 252 KASCADE Array detector stations were published. In our basic software framework of KCDC, KAOS, two new plugins handling the KCDC 'Publications' and first 'Spectra' data of related cosmic-ray experiments for download were included.
- A big change has been introduced with the release of **NABOO** in 2017. All data from KASCADE and KASCADE-Grande, recorded between 1997 and 2013 were published increasing the number of events to 433 million with roughly 800 data words per event. Furthermore, the matching CORSIKA simulations for KASCADE and GRANDE for six different high-energy hadronic interaction models were released.

- With the release **OCEANUS** in 2019 we moved the DataShop Mongo DB to a sharded cluster speeding up the processing time for user requests by roughly a factor of 50. Furthermore, the data of the Radio LOPES detectors were added as a new detector component within the KASCADE DataShop.
- With the release **PENTARUS** in 2020 we added a second DataShop, which allows now to add in a modular way for each new experimental data independent access on the same platform. This way we added a second DataShop with data sets based on KASCADE-Grande called COMBINED, handling the joint data analysis of the KASCADE and GRANDE detector arrays.
- With the release **SKARAGAN** published in 2021, we introduced data from the Armenian air-shower experiment Maket-Ani for the first time a DataShop not related to the KASCADE-Grande experiment.

In between these ‘major releases’ we published some ‘minor releases’ when, for example, new ‘Spectra’ or ’Simulations’ have been added or when bug-fixes were necessary. From the very beginning, a large interest from the community was given, proved, e.g. by our anonymous monitoring of the access to the portal with more than 400 users registered from five continents. We track page views and downloads with MATOMO analytics [5] to learn about the customers needs. As communication platform serves an Email list for the KCDC subscribers as well as social networks, like Twitter https://twitter.com/KCDC_KIT.

2. New in KCDC

2.1 Improving performance and stability

With the KCDC release **SKARAGAN 2.0** in 2021, we have updated all front-end and back-end software packages to the latest versions. KCDC is now running on the operating system UBUNTU 20.04 LTS. The outdated ftp-download procedure was switched to a direct https-download, which is valid for all downloads like user requests, pre-selections, simulations, programming examples etc. and we improved the privacy policy.

Early 2022, with the release **QUALOR 1.0**, we switched to a new mongoDB shard cluster, which increased the processing speed for the user requests by a factor of two. A new login procedure was added, which allows the users access to KCDC web portal and data via Helmholtz-AAI and for KIT internal users via a local Keycloak server.

2.2 Extending simulations and spectra data

Based on requests of our users, we have increased the number of available simulations by a factor of two. With release **SKARAGAN 2.0** we published a new simulation set based on the cosmic ray event simulation package CORSIKA using the high-energy interaction model SIBYLL 2.3d and Fluka for particle energies below 200 GeV.

In the folder ’Spectra’ at present more than 100 data sets are available for download from 27 different experiments published between 1984 and 2021 in the energy range 10^{14} to 10^{20} eV. The data sets contain published results, mostly from ground-based UHECR experiments. These are usually all particle spectra, but data sets from various mass groups like p, He, C, Si and Fe or heavy and light respectively derived from the unfolding procedure for different high-energy interaction models like QGSJet, EPOS and SIBYLL have also been published, such as for KASCADE-Grande.

2.3 Event Displays

In December 2022, with the release **QUALOR 2.0** we included a new plugin into the KCDC framework to visualise cosmic ray events measured and reconstructed by KASCADE-Grande. Now, within the KCDC web portal, we provide two possibilities to display events;

- *random events* from predefined displays (not from simulations);
- *generate your own event displays* for data from the KCDC DataShops and associated 'pre-selections' or 'simulations', using the software packages provided there.

We chose this way, which at first glance seems complicated, because there is no way for the KCDC users to access our event database directly.

2.3.1 Random events from predefined displays

From the entire data sample recorded with the KASCADE-Grande detector system, which consists of over 433 million events, we offer more than 210.000 events for random display. This selection represents a uniform distribution of the measured data in the energy range 10^{14} to 10^{18} eV. In case of KASCADE, about 123.000 events are available, while for GRANDE more than 88.000 events can be displayed.

Furthermore, to increase the number of events in the interesting high-energy region, we provide about 12.000 events for GRANDE in the region $N_{\mathrm{ch}} > 6.3$ [log10] and 10.000 events in case on KASCADE in the region $N_{\mathrm{e}} > 5.6$ [log10] (marked red in fig. 2).

For measured events, the event displays show 'Energy Deposits' and 'Arrival Times' recorded in each detector station of the KASCADE or GRANDE detectors.

In case of a KASCADE event, three plots are displayed, showing the energy deposits in the 252 e/γ-detectors of the KASCADE-Array (fig. 2, top left) and in the 192 μ-detectors (top right) in *eV*. The arrival time distribution is given in *ns* and represents the first time stamp of each detector station that has been hit by a charged particle.

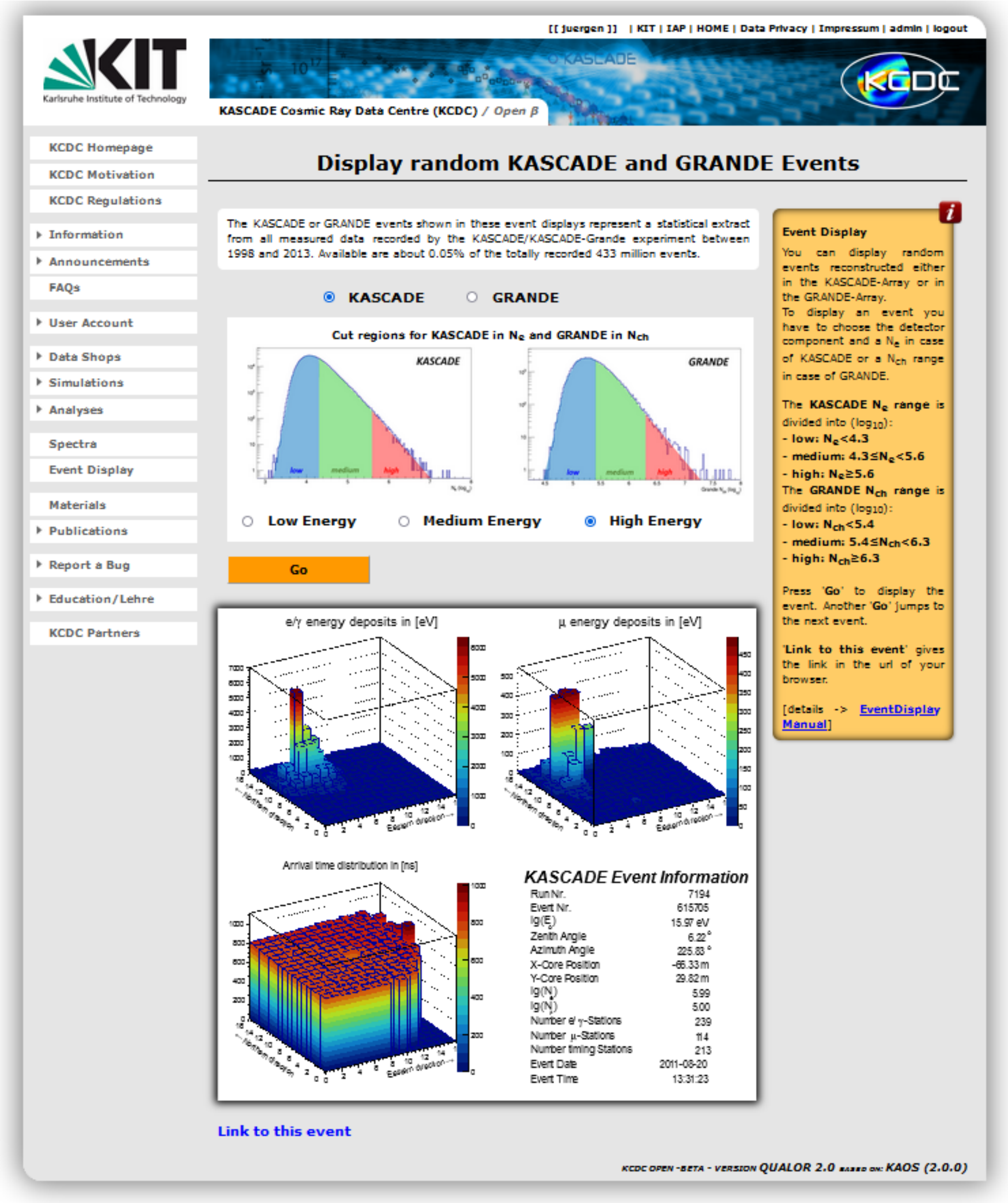
In case of a GRANDE event, the displays show the charged energy deposits in the 37 stations of the GRANDE-Array (top) in 'lego' and in 'colour' plots. The arrival time distribution in 37 detector stations (bottom left) is a relative time given in *ns* and represents the first time stamp of each detector station that has been hit by a charged particle.

A list of the event properties like run- and event numbers and event time, shower parameters reconstructed like core position and arrival direction as well as the respective number of stations contributing to this event for both detector components is shown in the bottom right corner.

2.3.2 User generated event displays

By means of C++ programs provided for download, the user can easily visualize Events measured by the KASCADE and GRANDE detectors as well as simulated events and pre-selections downloaded from the KCDC Web portal.

The Event Displays show 'Energy Deposit' and 'Arrival Time' distributions recorded in each detector station of the KASCADE or GRANDE detectors. In case of simulated events, the 'En-



Display random KASCADE and GRANDE Events

The KASCADE or GRANDE events shown in these event displays represent a statistical extract from all measured data recorded by the KASCADE/KASCADE-Grande experiment between 1998 and 2013. Available are about 0.05% of the totally recorded 433 million events.

● KASCADE ○ GRANDE

Cut regions for KASCADE in N_e and GRANDE in N_{ch}

○ Low Energy ○ Medium Energy ● High Energy

Event Display

You can display random events reconstructed either in the KASCADE-Array or in the GRANDE-Array.
To display an event you have to choose the detector component and a N_e in case of KASCADE or a N_{ch} range in case of GRANDE.

The **KASCADE N_e range** is divided into ($\log_{10}$):
- **low: N_e<4.3**
- **medium: 4.3≤N_e<5.6**
- **high: N_e≥5.6**

The **GRANDE N_{ch} range** is divided into ($\log_{10}$):
- **low: N_{ch}<5.4**
- **medium: 5.4≤N_{ch}<6.3**
- **high: N_{ch}≥6.3**

Press 'Go' to display the event. Another 'Go' jumps to the next event.

'Link to this event' gives the link in the url of your browser.

[details -> EventDisplay Manual]

e/γ energy deposits in [eV]
μ energy deposits in [eV]
Arrival time distribution in [ns]
Northern direction
Eastern direction

KASCADE Event Information

Run Nr.	7194
Event Nr.	615705
$\lg(E_0)$	15.97 eV
Zenith Angle	6.22 °
Azimuth Angle	225.83 °
X-Core Position	-66.33 m
Y-Core Position	29.82 m
$\lg(N_e)$	5.99
$\lg(N_\mu)$	5.00
Number e/ γ-Stations	239
Number μ-Stations	114
Number timing Stations	213
Event Date	2011-08-20
Event Time	13:31:23

Link to this event

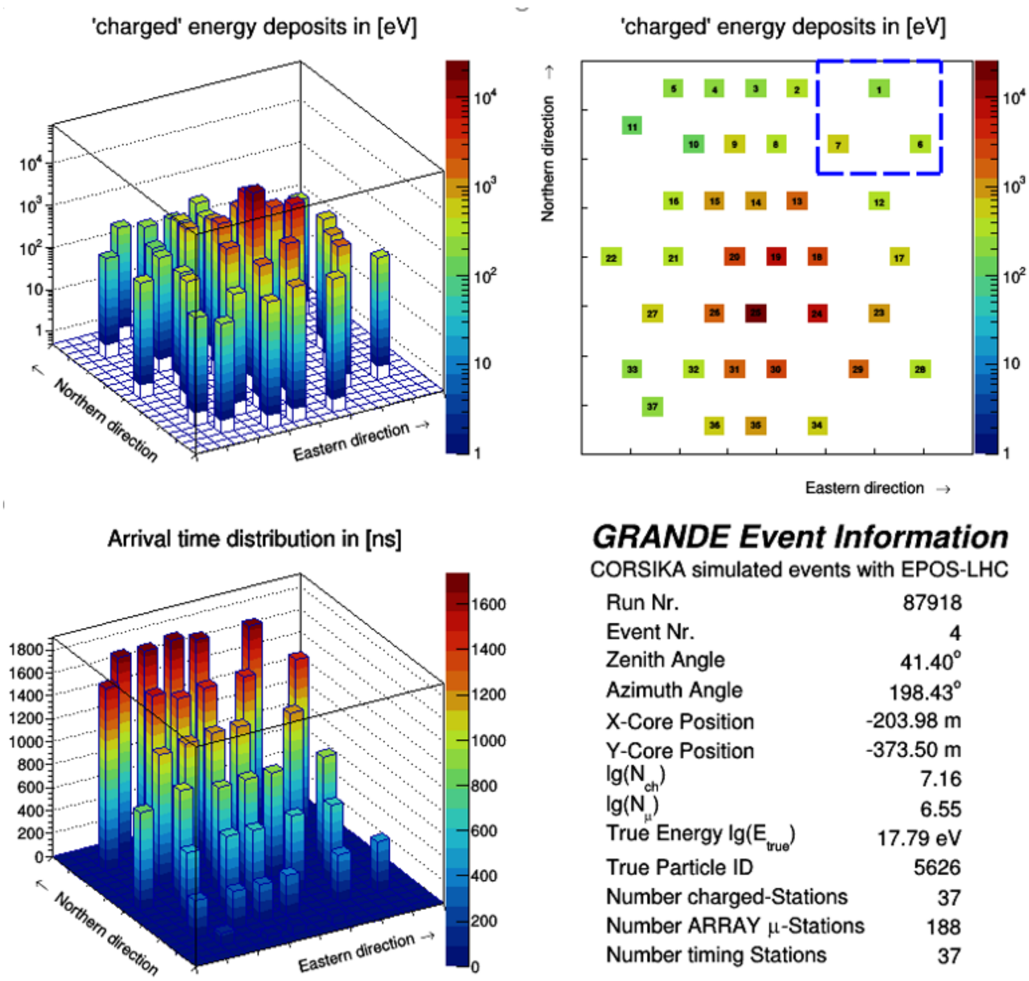
KCDC OPEN -BETA - VERSION QUALOR 2.0 BASED ON: KAOS (2.0.0)

Figure 2: Display of a KASCADE cosmic ray air shower event in KCDC

ergy Deposits' and 'Arrival Times' of air showers, simulated with the air shower simulation code CORSIKA[1] and the detector simulation code CRES[2] are displayed.

As the provided programs are embedded in the CERN ROOT framework, only ROOT files can be analysed and displayed. An example for a high-N_{ch} GRANDE event is given in fig. 3.

[1] COsmic Ray SImulation for KAscade
[2] Cosmic Ray Event Simulation



GRANDE Event Information

CORSIKA simulated events with EPOS-LHC

Run Nr.	87918
Event Nr.	4
Zenith Angle	41.40°
Azimuth Angle	198.43°
X-Core Position	-203.98 m
Y-Core Position	-373.50 m
$\lg(N_{ch})$	7.16
$\lg(N_{\mu})$	6.55
True Energy $\lg(E_{true})$	17.79 eV
True Particle ID	5626
Number charged-Stations	37
Number ARRAY μ-Stations	188
Number timing Stations	37

Figure 3: Display of a simulated GRANDE event in KCDC

2.4 Tutorials and Masterclasses at KCDC

The integration of Jupyter Notebooks support directly on the KCDC platform has allowed users to conduct data analysis in an online mode [4]. To take advantage of this capability, logged-in KCDC users can navigate to the 'Analyses' tab on the website's left panel and select 'Jupyterhub'. This will redirect them to the Jupyter server directly connected to KCDC. Access to the server is granted through the respective KCDC user account. The Jupyter environment serves as a versatile tool, facilitating not only individual work but also conduction of tutorials and masterclasses. Our first experience of organising a fully online master class (due to the prevailing Corona conditions) based on KCDC data using the Jupyter instance at KIT/IAP, was presented in [6].

Presently, the Jupyter Hub at IAP provides users with access to five tutorials based on the data by KASCADE and Tunka-Rex experiments, as well as open data sourced from the IceTop setup [7] of the IceCube experiment These tutorials are currently offered in English and/or German, with future plans to develop versions in both languages for all tutorials.

Notable attention should be given to an invited lecture-masterclass conducted at the German

DPG Spring Meeting of the Matter and Cosmos Section (SMuK) [8], as part of the Hacky Hours - an experimental joint session of the working groups "Young DPG" (AKjDPG) and the Working Group on Information (AGI). The session aimed to present and discuss practical tools for scientific work. In the first half of the event, we delivered a comprehensive report on KCDC within the context of open science. This included providing access to open scientific data and open educational resources, employing open-source software within the portal, exploring the handling of big data, and upholding the FAIR data principles in experimental particle astrophysics. Furthermore, we explained the structure and operational principles of the KCDC data center services. The second part of the event consisted of an online tutorial conducted on KCDC's JupyterHub platform. During the practical session, participants were invited to work with their own laptops and temporary user accounts provided by KCDC specifically for the tutorial.

The tutorial included the following steps:

1. Logging into the KCDC portal using the temporary account and familiarizing themselves with the available datasets and portal materials.
2. Transitioning to and logging into KCDC's JupyterHub.
3. Working with open data from the KASCADE experiment in Jupyter Notebook: Utilizing the KCDC API to load data; Reading the data into a pandas dataframe; Performing exploratory analysis on the data, including histogram visualization. In addition to the core tutorial, participants were presented with bonus tasks: (i) Comparing distributions of experimental data and simulation data, (ii) Developing a prototype of a lightweight data-driven application inspired by the development experience of the outreach machine-learning-based application by utilizing data from KCDC.

A total of 15 scientists, who were participants of DPG-2023, attended the tutorial.

3. Partnership with CRDB

Considering the vast amount of academic repositories and search engines for locating and accessing published scientific data, unified access to published datasets and spectra is still in the early stages. This is due to the large variety of experiments and thus the large variety (and it's differences in location, operation and data format) of measured data. In cooperation with CRDB, the *Cosmic-Ray DataBase*[3] [9, 10], KCDC is taking a step towards unification by embedding the spectra data from KCDC, i.e. published data from extensive air shower experiments, into CRDB. The advantage of such an extensive collection of UHECR data is that data from other experiments, even from balloon and satellite experiments, can be obtained relatively quickly via the CRDB web interface.

4. Outlook

Following requests of the community as well as the users of KCDC on how to handle big data in science, we have further developed KCDC, and have still several tasks and ideas on our to-do list.

[3] https://lpsc.in2p3.fr/crdb

An interesting cooperation is developing within the framework of the German initiative for a National Data Infrastructure, NFDI [11]. Here, the PUNCH4NFDI consortium [12] (representing the German communities in particle, astroparticle, nuclear and astro physics) is developing an overarching Science Data Platform that will make it possible to search for and combine digital research products from the various research areas. This is of particular interest for multi-messenger astroparticle physics. For PUNCH4NFDI, KCDC is an initial use-case to demonstrate the feasibility of the concept. Keep track on the KCDC platform, where we will inform you on next steps.

Acknowledgements: The authors acknowledge the KASCADE-Grande collaboration for their continuous support of the KCDC project. We are very grateful to CRDB people (in particular D. Maurin and H. Dembinski) for the cooperation with KCDC. In the PUNCH4NFDI consortium we thank A. Gaisser for his close cooperation. Our special thanks go to B. Keilhauer for her support on the administrative side, without whose help KCDC would not have progressed so quickly.

References

[1] A. Haungs et al.; 'The KASCADE Cosmic-ray Data Centre KCDC: Granting Open Access to Astroparticle Physics Research Data'; Eur. Phys. J. C (2018) **78**:741; https://doi.org/10.1140/epjc/s10052-018-6221-2

[2] H. Enke et al.; 'Survey of Open Data Concepts Within Fundamental Physics: An Initiative of the PUNCH4NFDI Consortium'; Comput.Softw.Big Sci. **6** (2022) 1, 6; https://doi.org/10.1007/s41781-022-00081-7

[3] M.D. Wilkinson et al.; 'The FAIR Guiding Principles for scientific data management and stewardship'; Scientific Data (2016) **3**: 160018; https://doi.org/10.1038/sdata.2016.18

[4] A. Haungs et al., 'Status and Future Prospects of the KASCADE Cosmic-ray Data Centre KCDC'; PoS ICRC2021 (2021), 422; https://doi.org/10.22323/1.395.0422

[5] MATOMO analytics; https://matomo.org/100-data-ownership/

[6] K. Link et al.; 'Online masterclass built on the KASCADE cosmic ray data centre'; PoS ICRC2021 1378; https://10.22323/1.395.1378

[7] IceTop-IceCube Masterclass. IceCube Collaboration, Accessed June 29, 2023. https://masterclass.icecube.wisc.edu/viewer/icetop.

[8] V. Tokareva at the DPG Spring Meeting of the Matter and Cosmos Section (SMuK). The German Physical Society (DPG), Accessed June 29, 2023. https://smuk23.dpg-tagungen.de/.

[9] D. Maurin et al.; 'A cosmic-ray database update: CRDB v4.1'; submitted 2023, e-Print: 2306.08901 [astro-ph.HE] https://doi.org/10.48550/arXiv.2306.08901

[10] D. Maurin et al.; 'Cosmic-Ray Database Update: Ultra-High Energy, Ultra-Heavy, and Antinuclei Cosmic-Ray Data (CRDB v4.0)'; Universe **6** (2020) 8, 102; https://doi.org/10.3390/universe6080102

[11] NFDI, Nationale Forschungsdateninfrastruktur https://www.nfdi.de/

[12] PUNCH4NFDI, Particles, Universe, Nuclei and Hadrons for the NFDI https://www.punch4nfdi.de/