
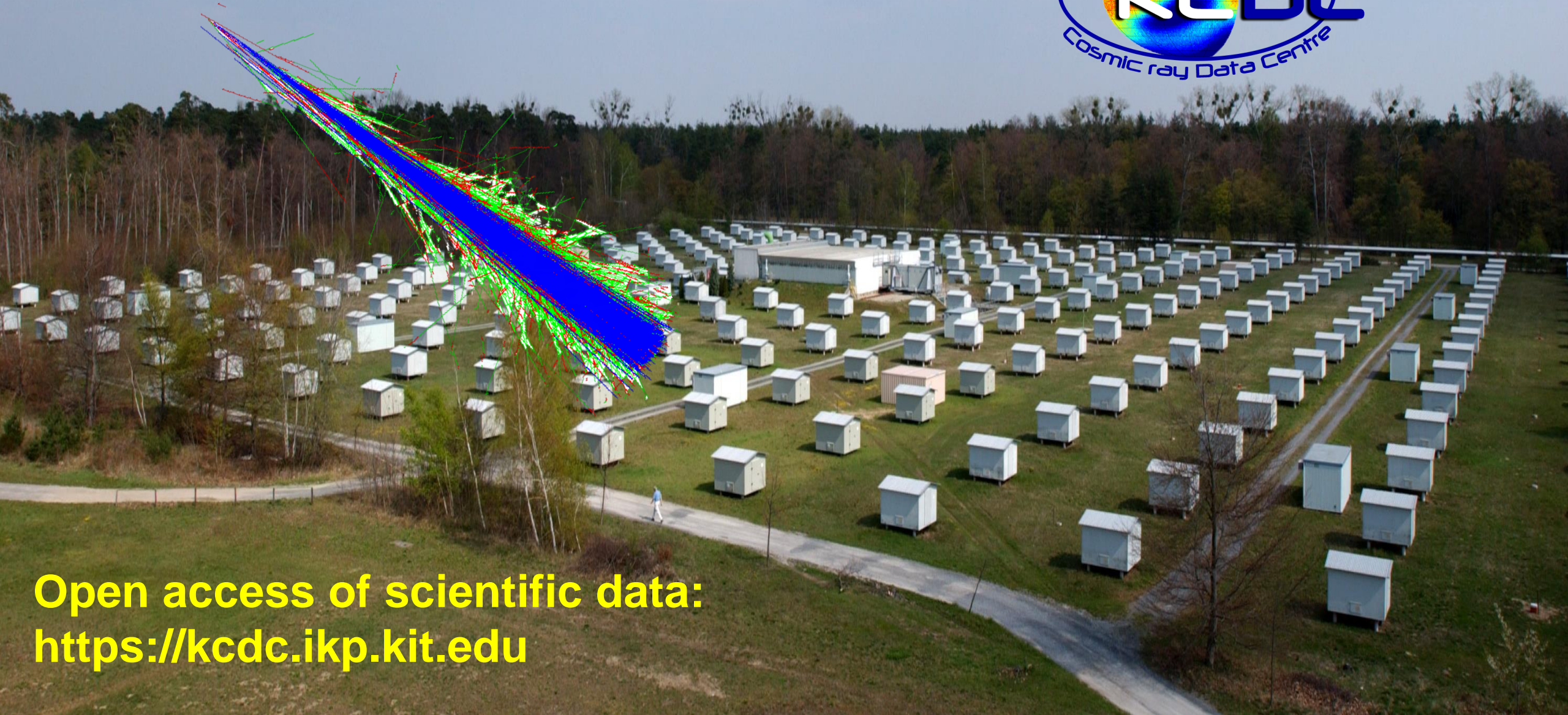


HAP workshop | Big Data Science in Astroparticle Physics

Aachen, 19-21 February 2018

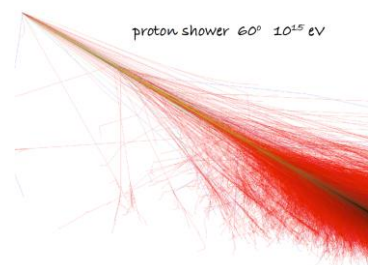
The logo for the KASCADE Cosmic ray Data Centre. It features a central circular graphic with a blue and yellow color scheme, possibly representing a cosmic ray shower or a celestial body. The word "KASCADE" is written in a stylized, bold, sans-serif font across the top of the circle. Below the circle, the words "Cosmic ray Data Centre" are written in a smaller, sans-serif font, following the curve of the bottom of the circle.



Open access of scientific data:
<https://kcdc.ikp.kit.edu>

KASCADE (-Grande) - timeline

- 58 collaborative papers in reviewed journals (4-5 still in queue, short author list papers not included)
- 57 PhD theses
- 86 diploma/master theses



Proposal

model tests

γ limit
light knee

(an)isotropy

proof-of-principle
radio detection

muon production height

EAS GHz emission
iron knee

light ankle

Xmax by radio

γ limit

1992

CORSIKA

1996

KASCADE

2000

KASCADE-Grande

LOPES



Cosmic Revelation

2004

Karlsruhe Air Shower
Test Facility

CROME

CROME

2008

2012

dismantling

KCDC

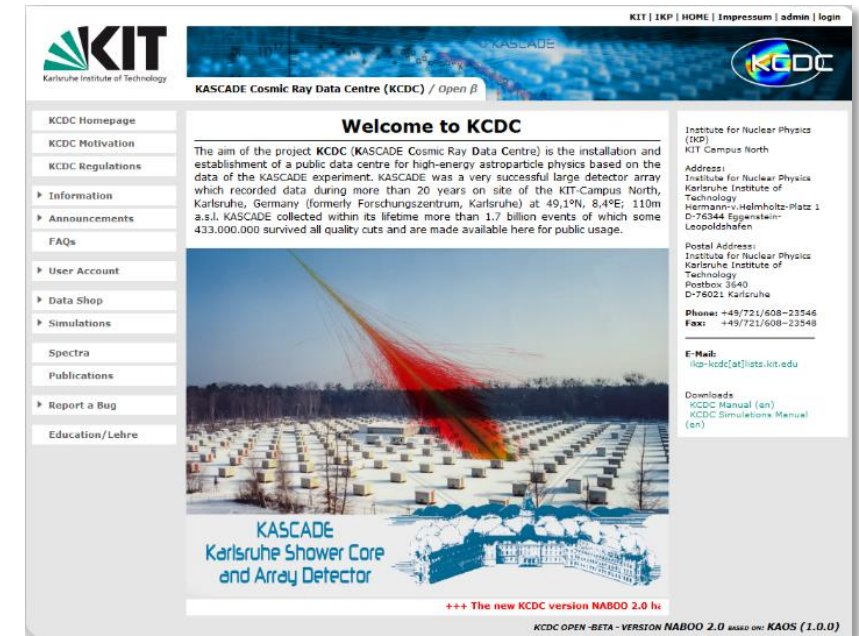


2017



KASCADE Cosmic ray Data Centre

- Motivation and Idea of Open Data:
 - public access to the data
 - data has to be preserved for future generations
- Web portal:
 - modern software solution
 - release the software as Open Source
 - educational courses
- Data access:
 - new release (Feb. 2017) with $4.3 \cdot 10^8$ EAS
 - simulation data
- Pioneering work in publishing research data in astroparticle physics



<https://kcdc.ikp.kit.edu/>

[J.Phys.Conf.Ser. 632 (2015) 012011]

KCDC in a nutshell

- providing open access to astroparticle physics research data as required by funding agencies



- **data provider**

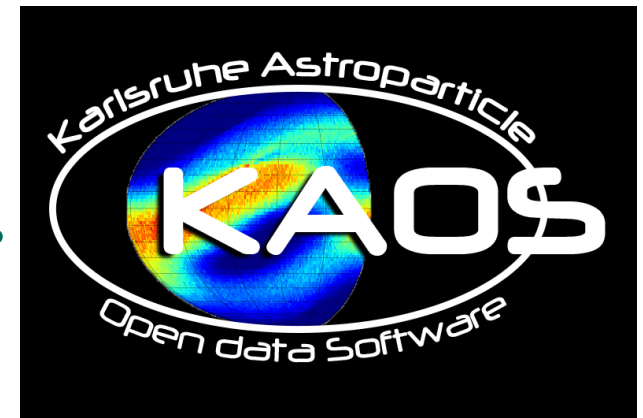
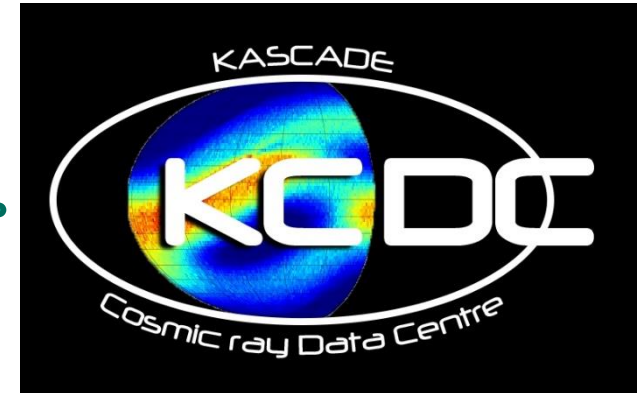
- follows the “Berlin Declaration on Open Data and Open Access”
- free, unlimited, open access to KASCADE cosmic ray data
- selection of fully calibrated quantities and detector signals
- reliable data source
- guaranteed data quality

- **information platform**

- experiment description
- meta information for data analysis
- physics background
- use of modern and open source web technologies
- tutorials (focused on teachers and pupils)

- **as long-term digital data archive**

- archive of software and data
- for the collaboration
- for the public



<https://kcdc.ikp.kit.edu/>

KCDC data shop

KIT Karlsruhe Institute of Technology

KASCADE Cosmic Ray Data Centre (KCDC) / Open β

[[haungs]] | KIT | IKP | HOME | Impressum | admin | logout

KCDC Data Shop

Components Available	Components Selected	Quantities and Cuts
Calorimeter	General Info	<input type="checkbox"/> Toggle all
	KASCADE	<input type="checkbox"/> Air Temperature range: -30 to 50 °C
	GRANDE	<input type="checkbox"/> Air Pressure range: 980 to 1040 hPa
		<input type="checkbox"/> DateTime range: 1998-05-08 to 2013-01-15
		<input type="checkbox"/> Global Time range: 8.945e+8 to 1.255e+9 sec
		<input type="checkbox"/> Mt range: 0 to 9.99e+8 m
		<input checked="" type="checkbox"/> Run Number range: 877 to 7417
		<input checked="" type="checkbox"/> Event Number range: 1 to 2e+6
		<input type="checkbox"/> e/y E-Deposit range: 0 to 2e+4 MeV
		<input type="checkbox"/> u E-Deposit range: 0 to 1000 MeV
		<input type="checkbox"/> Arrival Times range: -1550 to 2550 ns
		<input type="checkbox"/> Grande Deposit range: 0 to 1e+5 MeV
		<input type="checkbox"/> Grande Arrival range: 1000 to 1e+4 ns

Verify & Submit Request

Welcome to the Datasshop

On the left hand site you may select available detector components. Hovering the mouse over such a component, will give you some information on it. Once selected, you may click on the components name to view and select quantities associated with that detector. You may also deselect components using the left arrow. In right most column, you may select quantities for shipping and add cuts, that will be used to select only events passing these. On verification: Yellow means it has been corrected, red means you have to adjust your input. Is it a valid number? Is the lower bound larger than the upper bound?

[details -> KCDC Manual]

KCDC OPEN BETA - VERSION NAB00.00 based on KAOS (1.0.0)

Output:

zip-archive with data, metadata, and the EULA (end user licence agreement)

Data as ASCII, ROOT and HDF5 files

Commented header give information about the content

Tutorials and Teaching

- The goal: Providing the data to a general public

- Education portal

- first tutorials are up
(in German and English at the moment)
- knowledge database on KASCADE, astrophysics and related topics
- step by step tutorials of simple data analyses
- including a modern programming language code example
- interpretation and discussion of the outcome
- cooperation with local teachers and pupils
- later offering to teachers dedicated lessons for high schools

- introduction
- physics background
- step-by-step analysis
- source code example
- discussion
- interpretation
- pdf download of all

KIT
Karlsruhe Institute of Technology

KASCADE Cosmic Ray Data Centre (KCDC) / Open β

Access for teachers and pupils

zur deutschen Version

This is a compilation of interesting lessons within the vast field of cosmic radiation to illustrate the processes within and outside of our atmosphere by means of the data sets of the KASCADE experiment. This collection will be extended in cooperation with interested teachers and pupils to increase the understanding of the cosmic radiation.

The colors of the frames indicate the rating of the exercise:
red means 'heavy stuff', yellow denotes 'medium' and green 'rather easy' while blue can be considered as a 'finger exercise'.

How heavy is a cosmic particle?

Elektron-Muon-Verteilung

The cosmic radiation consists of atomic nuclei of positive charge reaching from hydrogen (1 proton) up to Iron (26 protons) travelling through space nearly at the speed of light and hitting the earth by chance. When entering the earth's atmosphere they collide with the molecules of the air and generate a variety of particles (mainly muons and electrons) which initiate collisions etc. This so called shower cascade can be detected with highly sophisticated instruments called detectors. These shower measurements enable us to determine the properties of the primary cosmic particle like its mass and energy.

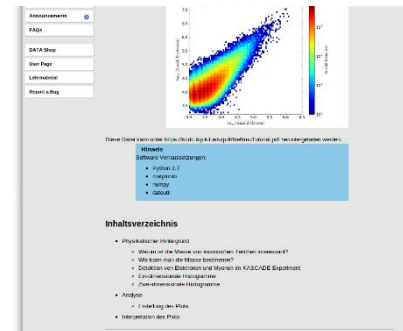
here you get: [exercise](#) - [tutorial](#) - [solution](#)

How does KASCADE see the sky?

Earth bound detector systems have usually a very narrow field of view of the sky. This applies for telescopes and for detectors of cosmic radiation likewise. To show the visible part of the sky for each experiment the celestial coordinates have to be determined from the arrival direction and the arrival time of each cosmic shower and plotted in a so called 'skyplot'. The skyplot shown here illustrates the frequency of cosmic showers from certain directions as measured by KASCADE. In a further step in the analysis chain the angular bins could be weighted with the exposure time.

here you get: [exercise](#) - [tutorial](#) - [solution](#)

KCDC OPEN DATA - VERSION NABOO.00 based on: KAOS (1.0.0)

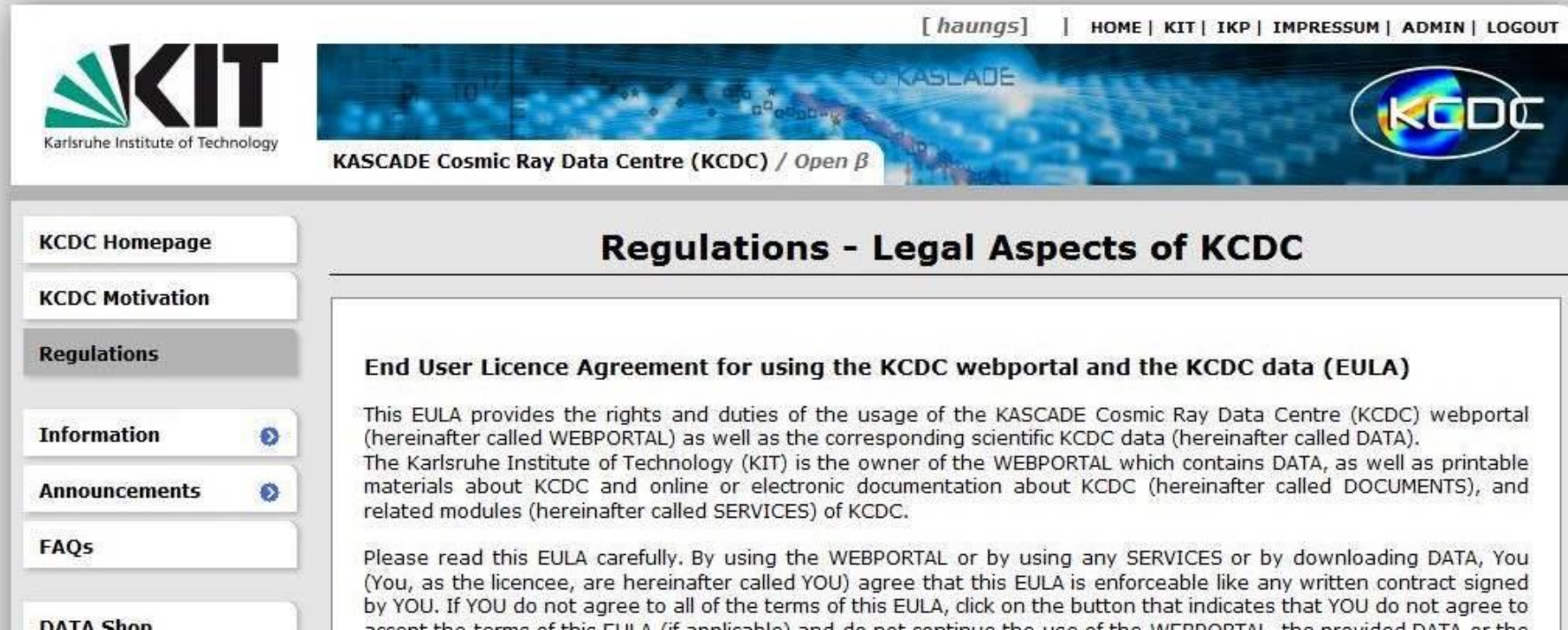


open data publication

- no ready available open data licence
- free access to data and web portal
- good scientific practice for work with data
- citation of collaboration, KIT, and web portal mandatory
- free redistribution of data “as is”

KCDC approach


- licence based on EULA model (as usually for software)
- licence details: following the industry
- flexible and adaptable to our needs
- signed during registration
- shipped with each data package




NABOO 2.0 is released!

27.10.2017





KIT
Karlsruhe Institute of Technology



KASCADE Cosmic Ray Data Centre (KCDC) / Open β

KCDC Homepage

KCDC Motivation

KCDC Regulations

Information

Announcements

FAQs

User Account

Data Shop

Simulations

General Info

QGSjet-II-02

QGSjet-II-04

EPOS-1.99

EPOS-LHC

SIBYLL-2.1

SIBYLL-2.3

Spectra

Publications

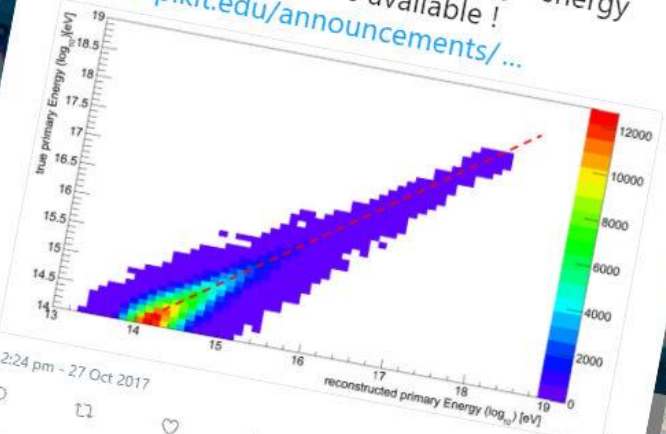
Report a Bug

Education/Lehre

Welcome to KCDC

The aim of the project **KCDC** (KASCADE Cosmic Ray Data Centre) is the installation and establishment of a data centre for high-energy astroparticle physics based on the data of the KASCADE experiment. The data were collected during the last decade by a very successful large detector array which recorded cosmic ray events at the Karlsruhe Institute of Technology (KIT) Forschungszentrum für Experimentelle Kernphysik. The data are collected within its lifetime more than 10¹² events and are made available to the scientific community.

#KCDCversionNaboo 2.0 is released!
Simulation sets for 6 different high energy simulation models are available!
[kcdc.ikp.kit.edu/announcements/...](http://kcdc.ikp.kit.edu/announcements/)



12:24 pm - 27 Oct 2017

Tweet your reply

Karlsruhe Snow and Array Detector

+++ The new KCDC version NABOO 2.0 has been released !!! +++

Institute for Nuclear Physics (IKP)
KIT Campus North

Address:
Institute for Nuclear Physics
Karlsruhe Institute of Technology
Hermann-v.-Helmholtz-Platz 1
D-76344 Eggenstein-Leopoldshafen

Postal Address:
Institute for Nuclear Physics
Karlsruhe Institute of Technology
Postbox 3640
D-76021 Karlsruhe

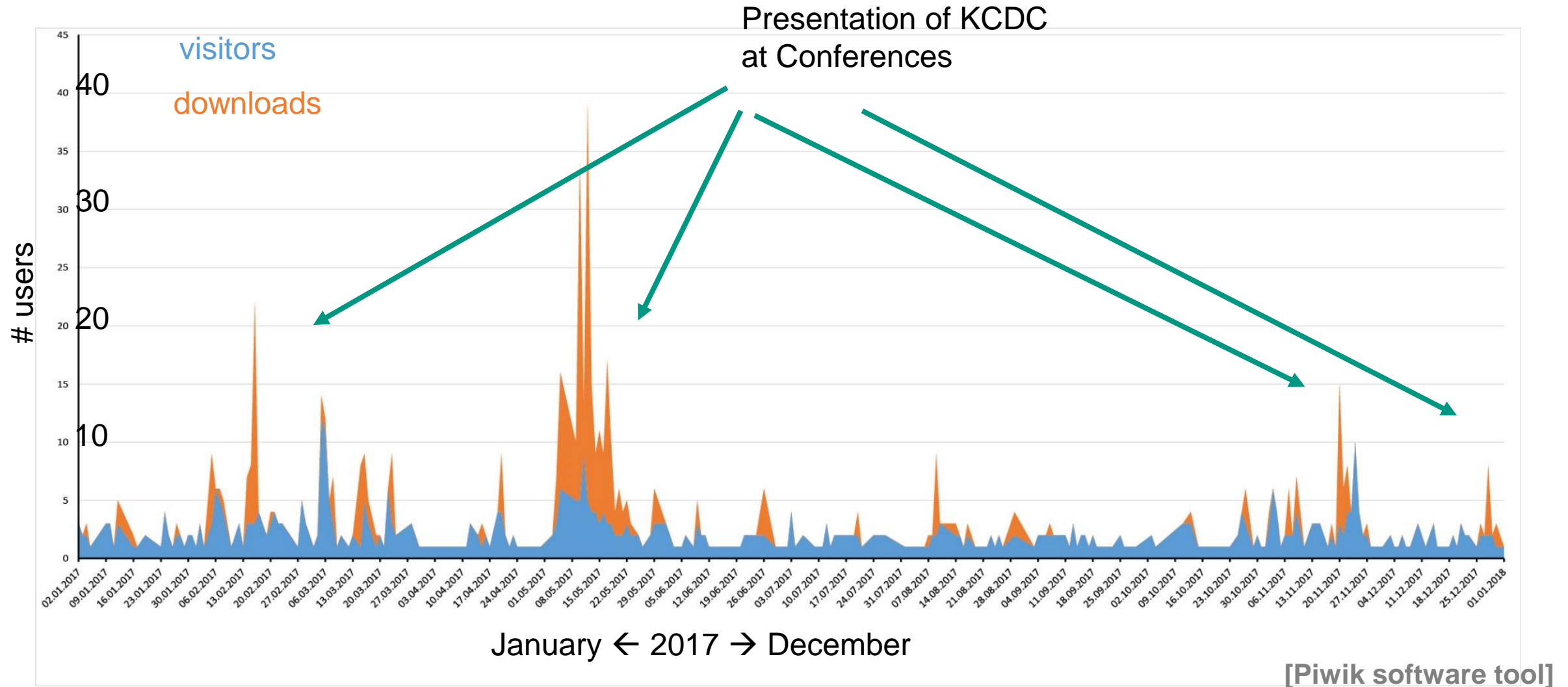
Phone: +49/721/608-23546
Fax: +49/721/608-23548

E-Mail:
[ikp-kcdc\[at\]lists.kit.edu](mailto:ikp-kcdc[at]lists.kit.edu)

Downloads
[KCDC Manual \(en\)](#)
[KCDC Simulations Manual \(en\)](#)

KCDC OPEN -BETA - VERSION NABOO 2.0 BASED ON: KAOS (1.0.0)

KASCADE Cosmic ray Data Centre



Recommendations of the KAT (white paper)

„Astroparticle Physics in the Light of the Digitalen Agenda der Bundesregierung“*

Recommendations of the KAT

The KAT emphatically emphasises the importance of setting up and developing centres for data storage, the provision of data and the necessary computing resources as a basic digital service for German scientists and, moreover, for public participation in scientific data.

The KAT supports the establishment of a structure that facilitates communication between scientists as users of scientific data and modern data analysis methods on the one hand, and continues to implement expert advice within the framework of user support.

The KAT draws attention to the central importance of externally funded and sustainably invested human resources positions, which are absolutely necessary for the support of users.

* <https://www.bmbf.de/de/die-digitale-agenda-relevant-auch-fuer-bildung-wissenschaft-und-forschung-206.html>

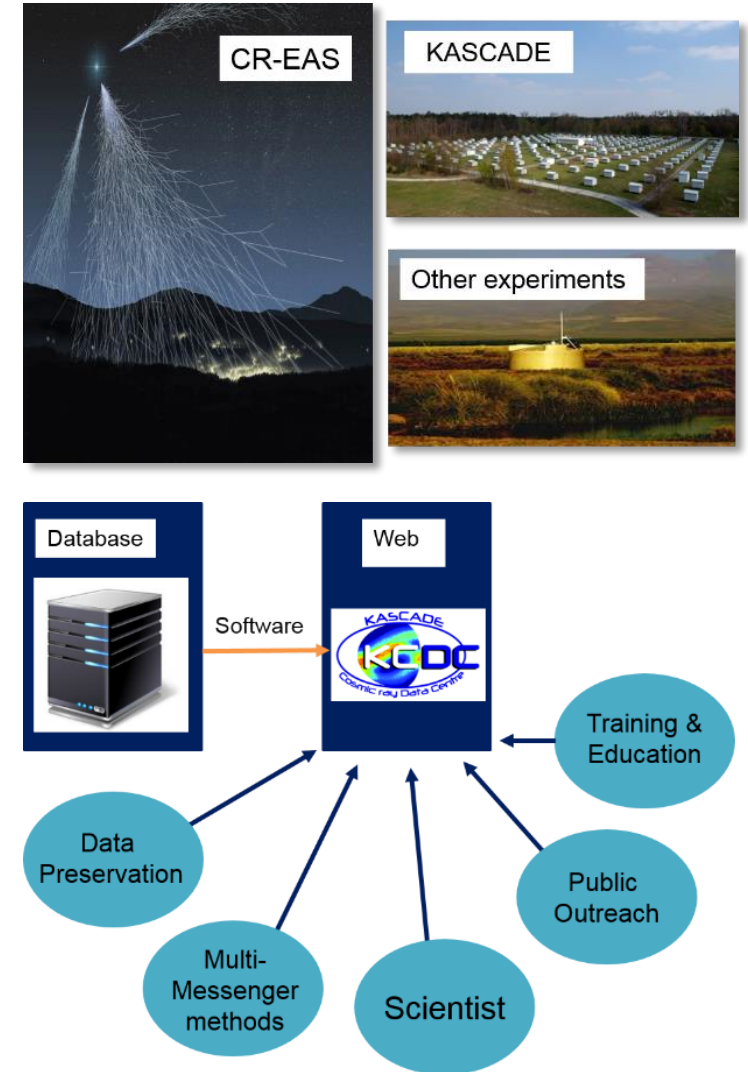
Towards a (global) Analysis & Data Centre in APP

Motivation:

- Astroparticle Physics requests for multi-messenger analyses.
This needs an experiment-overarching platform
- High demand in (German and international) community
- APP Observatories are globally distributed (no CERN or ESA)

Important steps:

- Develop an open science system based on KCDC and the KIT GridKa environment
- Develop integrated solutions of distributed data storage algorithms and techniques
- Allowing community to perform multi-messenger analyses with deep learning methods



Initiative for a (global) Analysis & Data Center in Astroparticle Physics

- Astroparticle Physics requests for multi-messenger analyses - this needs an **experiment-overarching** platform!

■ Tasks

- Provide sustainable access to scientific data
- Archiving of Data and Meta-Data
- Providing analysis tools
- Education in Big Data Science
- Development area for multi-messenger analyses (e.g. Deep Learning)
- Platform for communication and exchange within Astroparticle Physics

■ Elements

- Advancement, generalization of existing structures (like KCDC and others)
- In direction of a virtual Observatory (like in astronomy)
- In direction of Tier-systems and DPHEP (like in particle physics)
- „Digitale Agenda der Bundesregierung“
- OECD Principles and Guidelines for Access to Research Data from Public Funding
- Follow the FAIR principles of data handling

FINDABLE-ACCESSIBLE-INTEROPERABLE-REUSABLE



Analysis and Data Centre in Astroparticle Physics

Data availability

➤ Data availability:

All researchers of the individual experiments or facilities require quick and easy access to the relevant data.

➤ Analysis:

Fast access to the generally distributed data from measurements and simulations is required. Corresponding computing capacities should also be available.

➤ Simulations and methods development:

The researchers need an environment for the production of relevant simulations and the development of new methods (machine learning).

Simulations
& Methods
development

Open
access

➤ Open access:

More and more it is necessary to make the scientific data available not only to the internal research community, but also to the interested public: public data for public money!

➤ Education in data science:

Not only data analysis itself, but also the efficient use of central data and computing infrastructures requires special training.

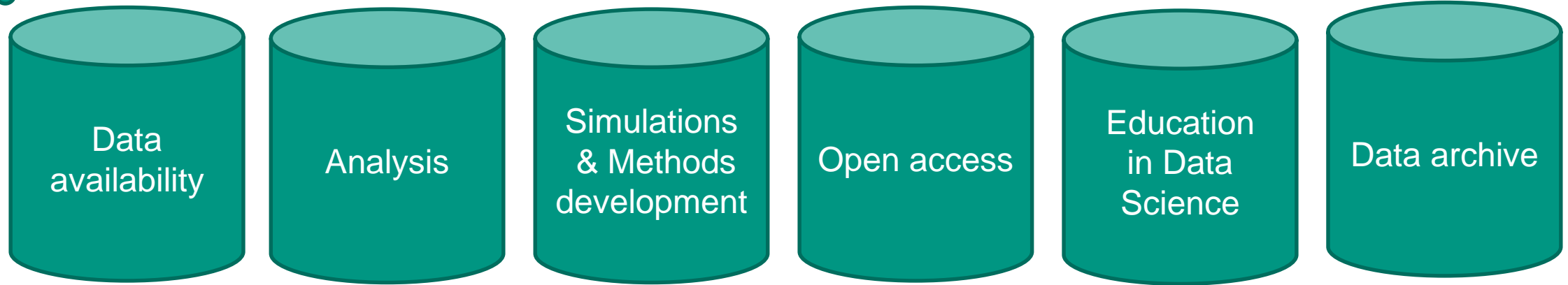
➤ Data archive:

The valuable scientific data and metadata must be preserved and remain interpretable for later use (data preservation).

Education
in Data
Science

Data archive

Analysis and Data Centre in Astroparticle Physics



- **Data preservation** ----
like DPHEP, KCDC
- **Metadata preservation** ----
like KCDC
- **Data storage (archive)** ----
like DPHEP, GridKa
- **Computing services (Grid vs. Cloud)** ---
like CERN Tier-centres
- **Data access (policy, technology, rate)** ---
like GridKa, KCDC
- **Training on Data use (maintenance, tutorials)** ---
like KCDC, VISPA, CDS
- **Data analysis, Simulation, modeling** ---
like GridKa, advanced VISPA?
- **Data science, workflows** (tools, e.g. deep learning, tutorials) ---
like VISPA
- **Data publication / Outreach** ---
like KCDC, masterclasses
- **Data education** ---
like KCDC, GridKa-school
- **Data exchange** ---
like AMON, GAVO
- **Data catalogues** ---
like Re3Data

Partly realized
in individual
experiments

Initiative for a (global) Analysis & Data Centre in Astroparticle Physics

National Data Centre for Astroparticle Physics

Initiative for a Data and Analysis Centre for Astroparticle Physics

2 November 2017
Karlsruhe Institute of Technology (KIT)
Europe/Berlin timezone

Overview

Scientific Programme

Timetable

Contribution List

Author List

Registration

Registration Form

Participant List



Campus Plan

for the slides, please click left on "Contribution List"!!!

preliminary Agenda: click left on 'Timetable' or below on 'Poster'

Organizing Committee: Andreas Haungs (KIT), Christian Stegmann (DESY), Achim Streit (KIT), Sabine Bucher (KIT)

November 2017:
40 Participants
Helmholtz + University groups

Goals:

- Data catalogues & computing resources
- Analysis & simulations
- User support & user platforms
- FAIR principles of data handling
**FINDABLE-
ACCESSIBLE-
INTEROPERABLE-
REUSABLE**
- Drafting white paper....

Karlsruhe-Russian Astroparticle Data Life Cycle Initiative

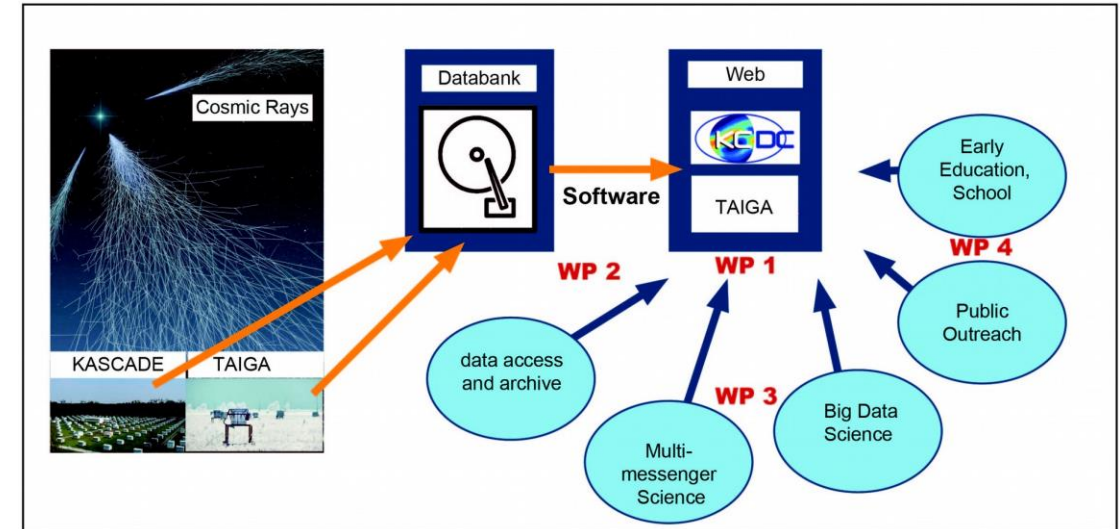
• Basics

- project period (2018-2020)
- Russia: SINP MSU, ISU, ISDCT SB RAS
Germany: KIT, DESY
- Team leaders: A. Kryukov (SINP MSU)
and A. Haungs + A. Streit (KIT)

• Main targets of the Project

- Extension example: data from Tunka/TAIGA and KASCADE-Grande
- Developing integrated solutions of distributed data storage techniques with a common meta-catalog
- Development of appropriate machine-learning techniques
- Perform experiment overarching multi-messenger astroparticle physics
- Learn to use GridKa environment
- Creation of an educational subsystem

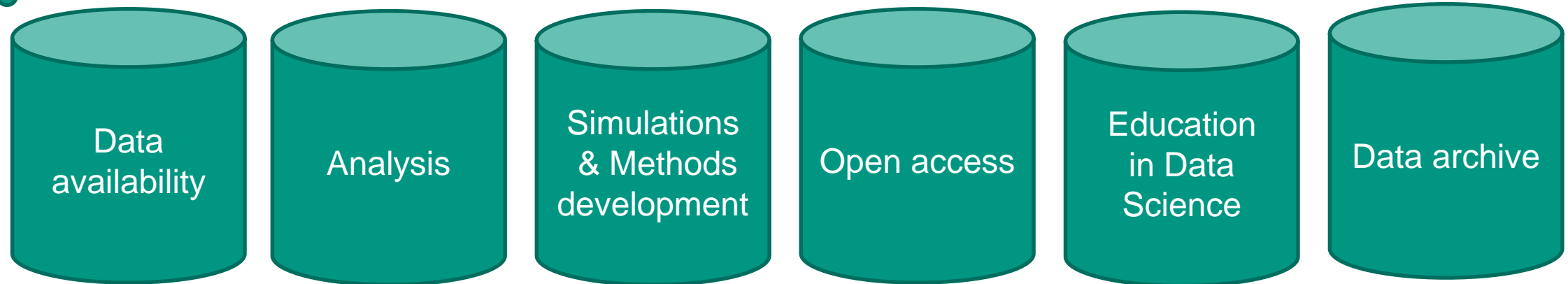
Open PhD positions!
Contact me.... ☺



**KRAD is a first step in extension
and generalization of KCDC**

Initiative for a (global) Analysis & Data Centre in Astroparticle Physics

Analysis and Data Centre in Astroparticle Physics



Next steps:

- Helmholtz & Universities define the specific needs.
- Secure funding & 'organize' hardware
- Implementation and: Start 😊