Karlsruhe Institute of Technology

HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

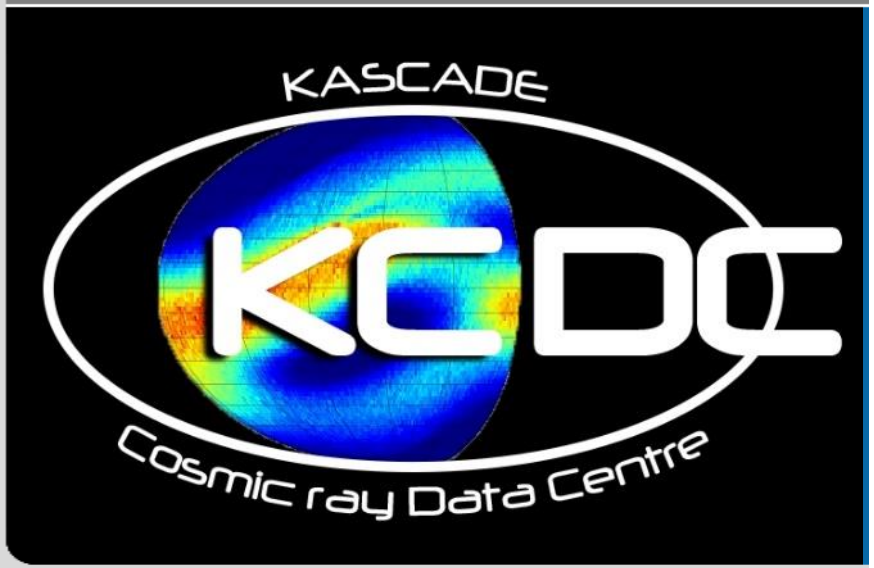
HELMHOLTZ | GEMEINSCHAFT
Allianz für Astroteilchenphysik

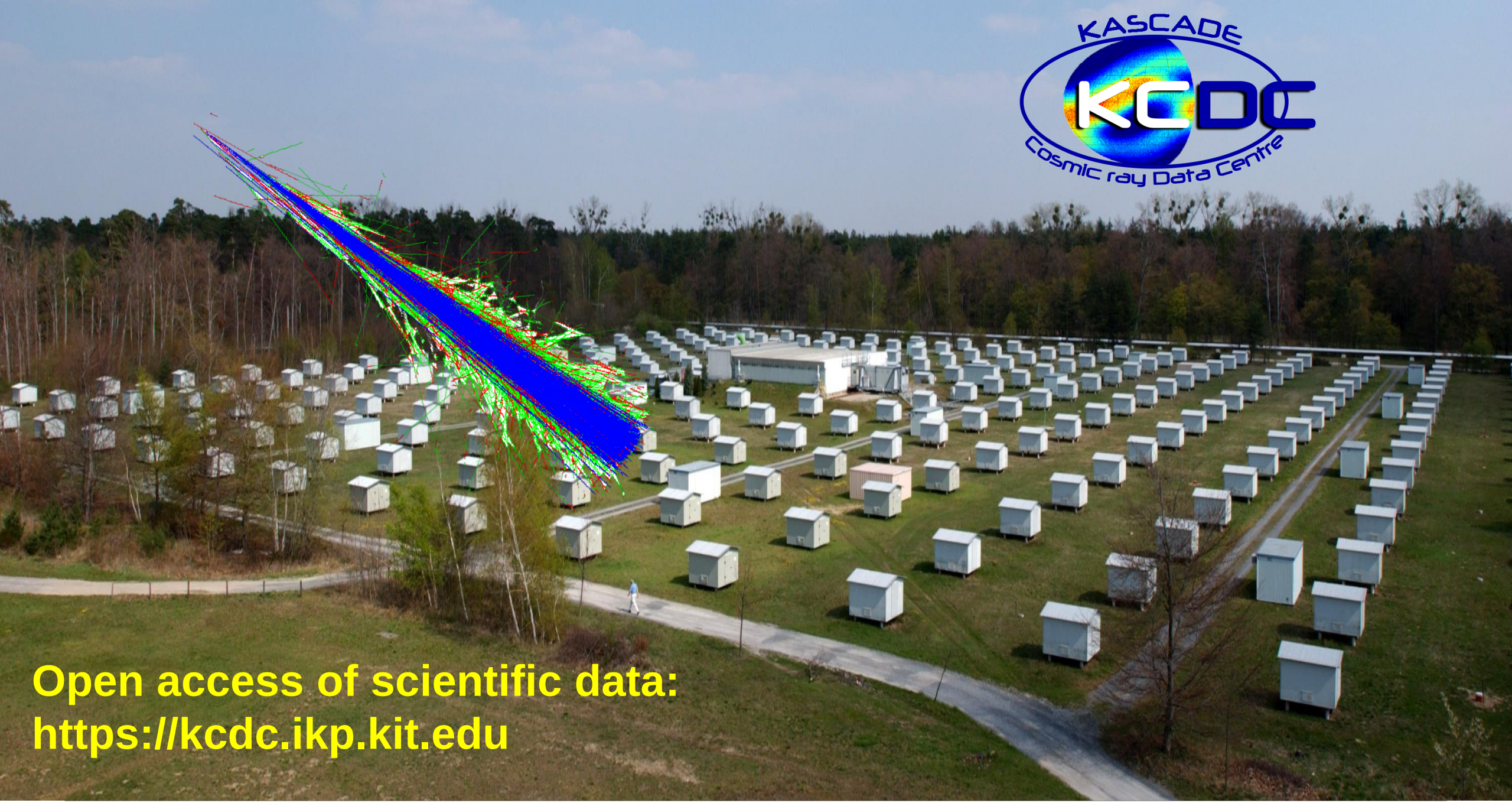
KAT. Komitee für Astro. Teilchen. Physik

Progress and Prospects of an Analysis and Data Centre in Astroparticle Physics

HAP workshop | Big Data Science in Astroparticle Physics
Aachen, 19-21 February 2018

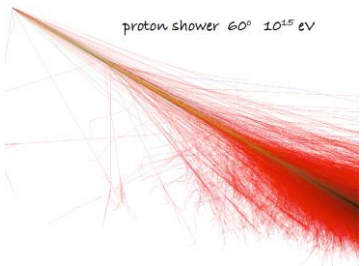
Andreas Haungs




KASCADE
KCDC
Cosmic ray Data Centre
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**



1992 Proposal
CORSIKA
1996 model tests
KASCADE
2000 light knee
γ limit
2004 (an)isotropy
LOPES
KASCADE-Grande
2008 proof-of-principle radio detection
muon production height
2012 iron knee
EAS GHz emission
light ankle
Xmax by radio
γ limit
2017 dismantling
CROME
Karlsruhe Air Shower Test Facility
Cosmic Revelation
KCDC

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES
KAT. Komitee für Astro. Teilchen. Physik

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

#KCDC_KIT

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

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

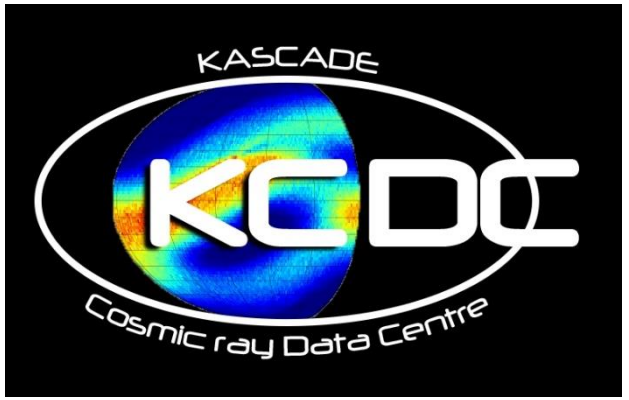
KCDC in a nutshell

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

#KCDC_KIT

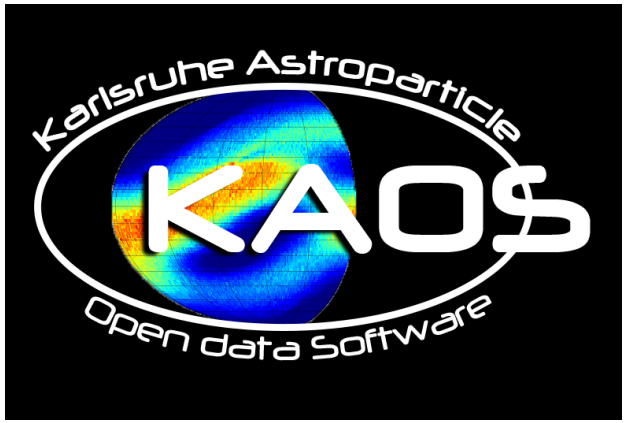
- **data provider**
  - follows the “Berlin Declaration on Open Data and Open Access”
  - free, unlimited, open access to KASCADE cosmic ray data
  - selection of fully calibrated quantities and detector signals
  - reliable data source
  - guaranteed data quality

The web portal



- **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)

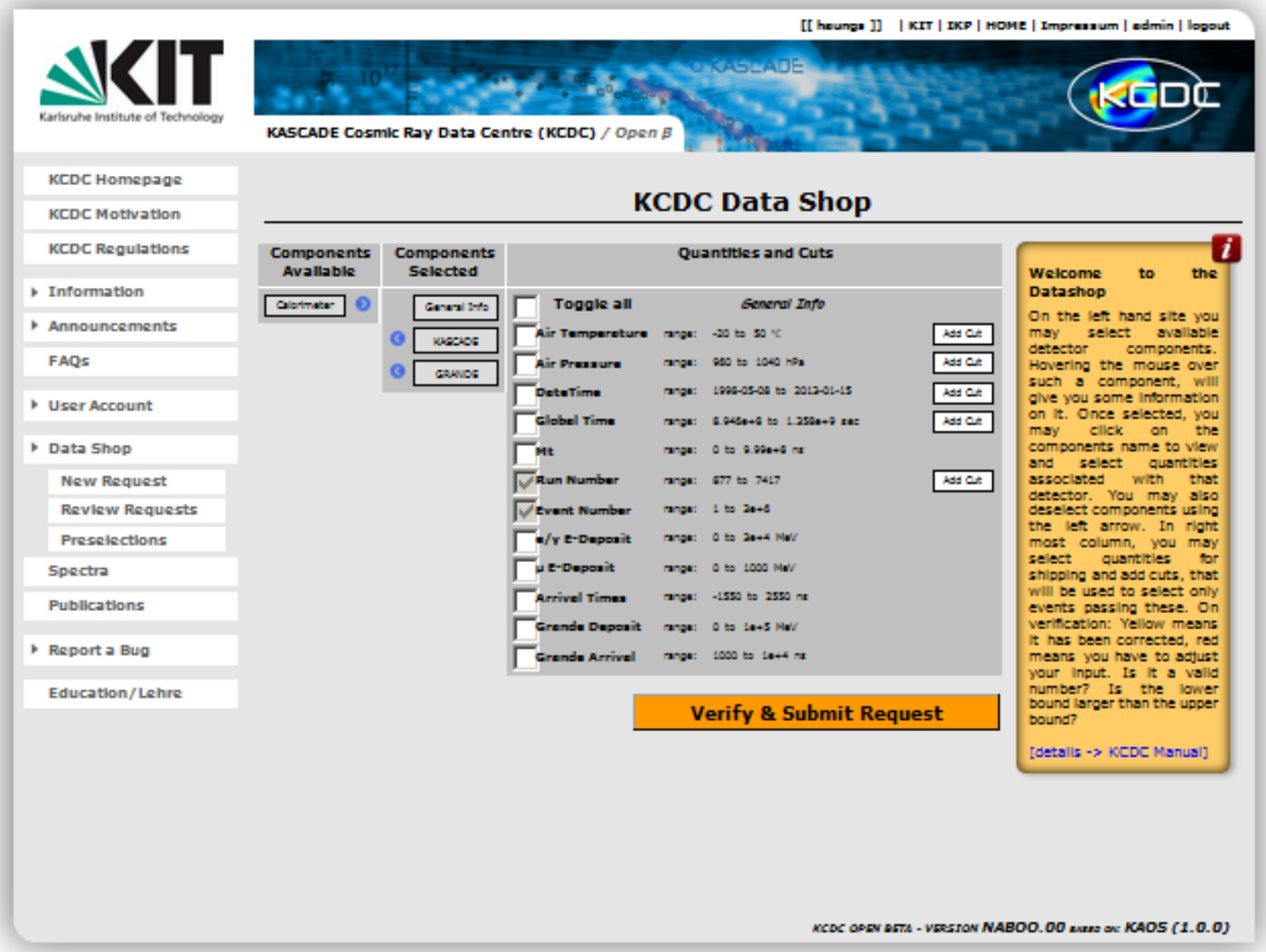
KCDC software



- **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



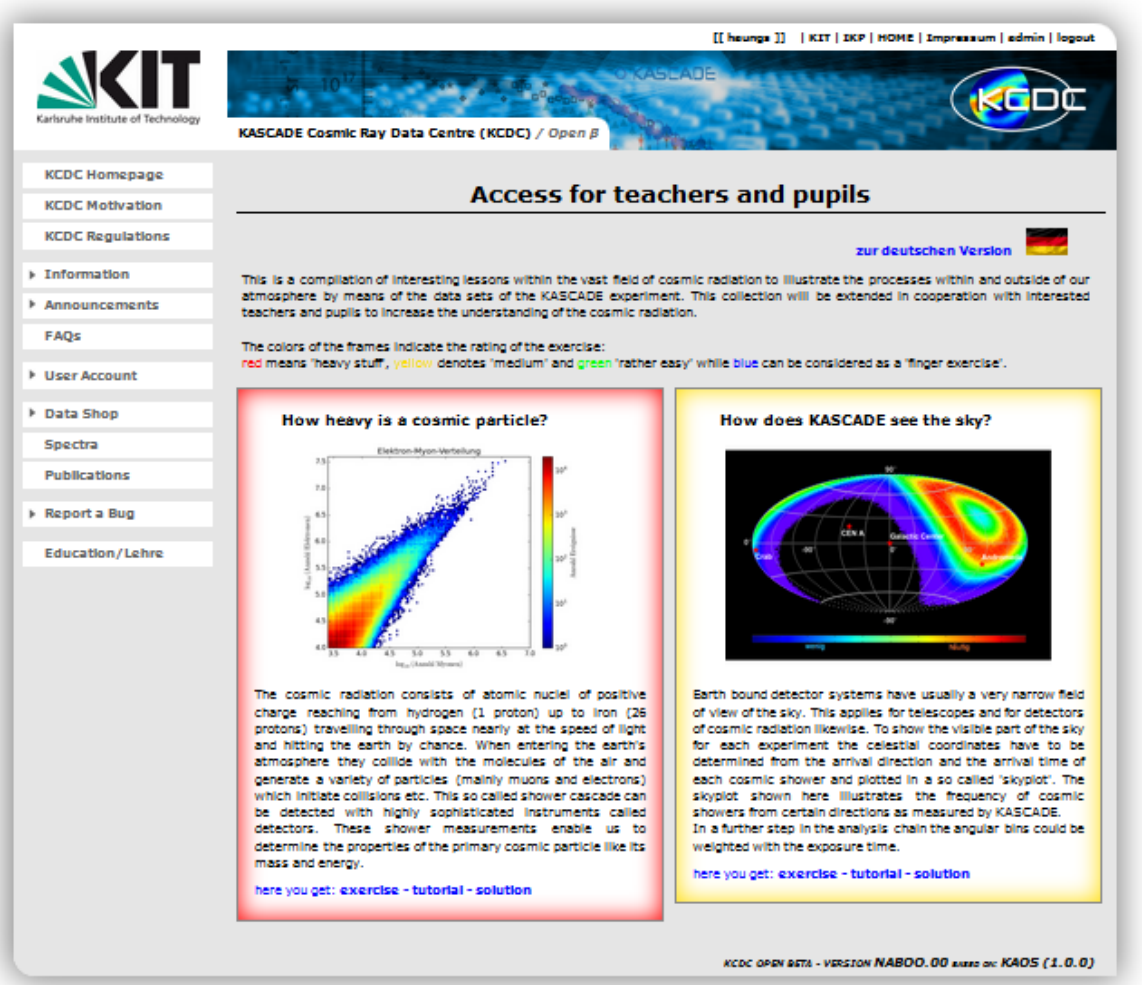
Output:

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

Data as ASCII, ROOT and HDF5 files

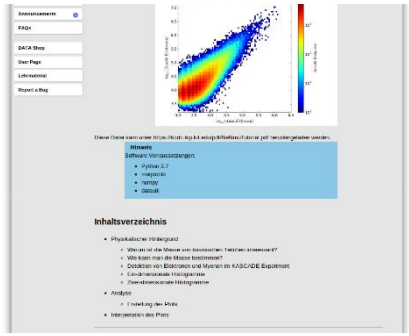
Commented header give information about the content

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

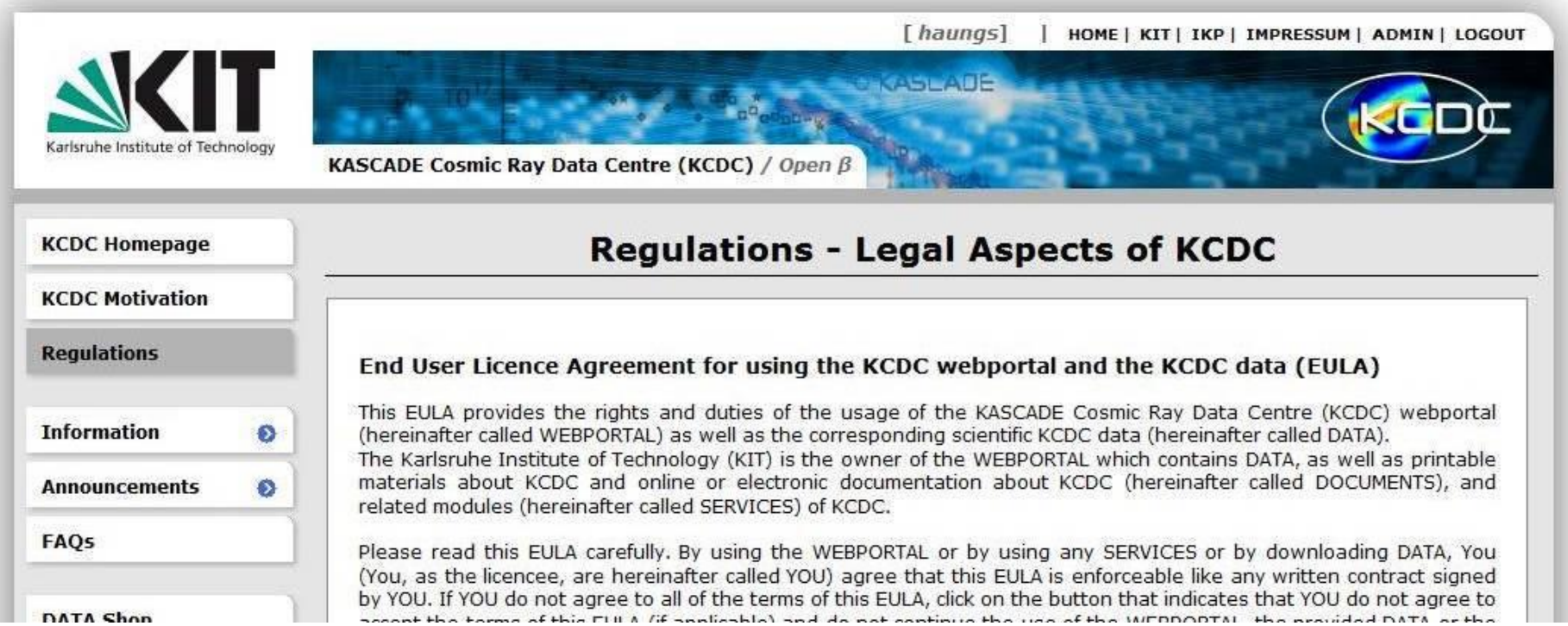
Law and Order

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”

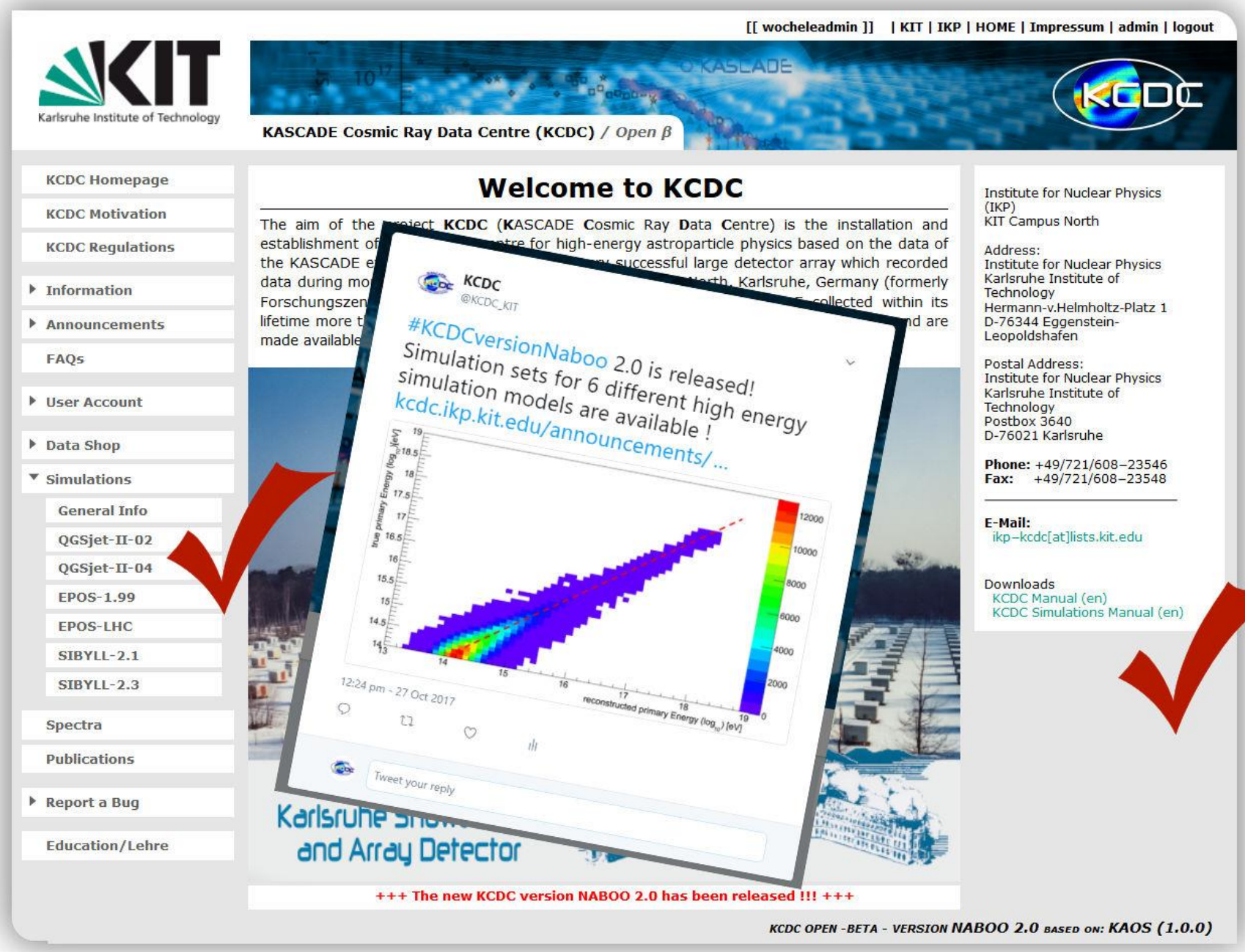
KCDC approach

- 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

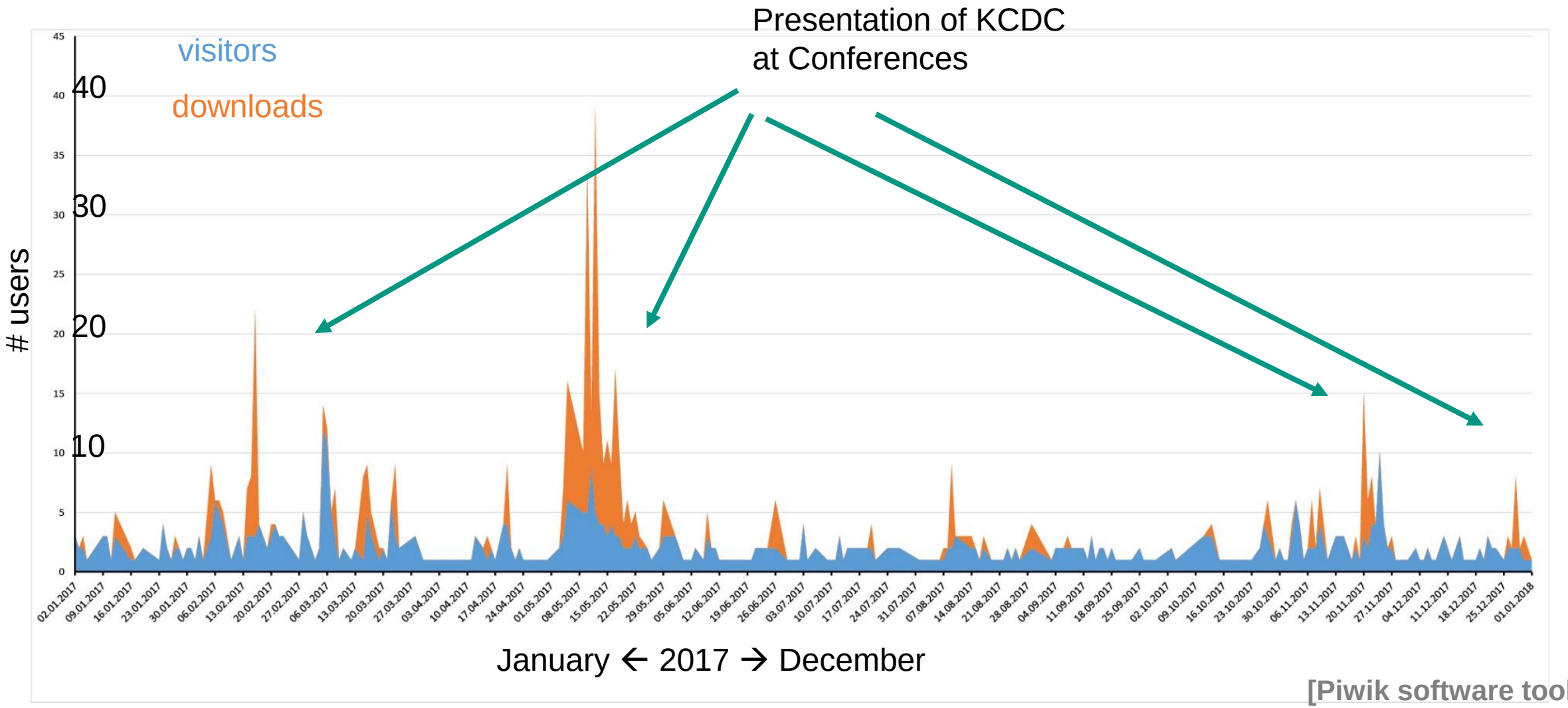


KASCADE KCDC Cosmic Ray Data Centre

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

KA Astro.Te

KASCADE Cosmic ray Data Centre



[Piwik software tool]

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

KAT. Komitee für Astro. Teilchen. Physik

KIT Karlsruhe Institute of Technology

Recommendations of the KAT (white paper)

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

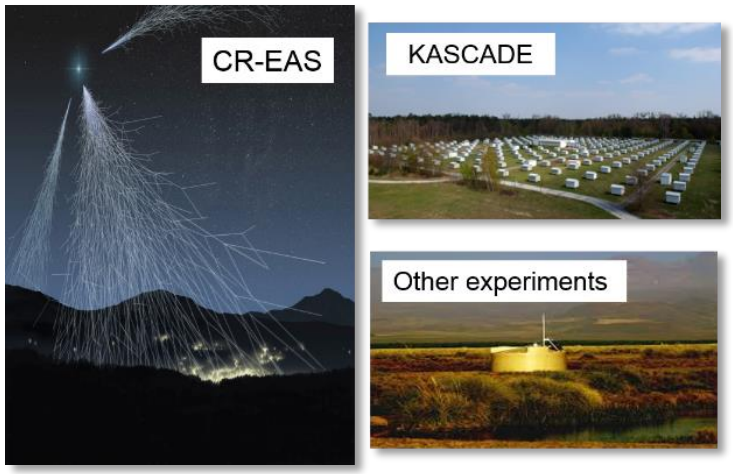
Recommendations of the KAT
The KAT emphatically emphasises the importance of setting up and developing centres for data storage, the provision of data and the necessary computing resources as a basic digital service for German scientists and, moreover, for public participation in scientific data.
The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

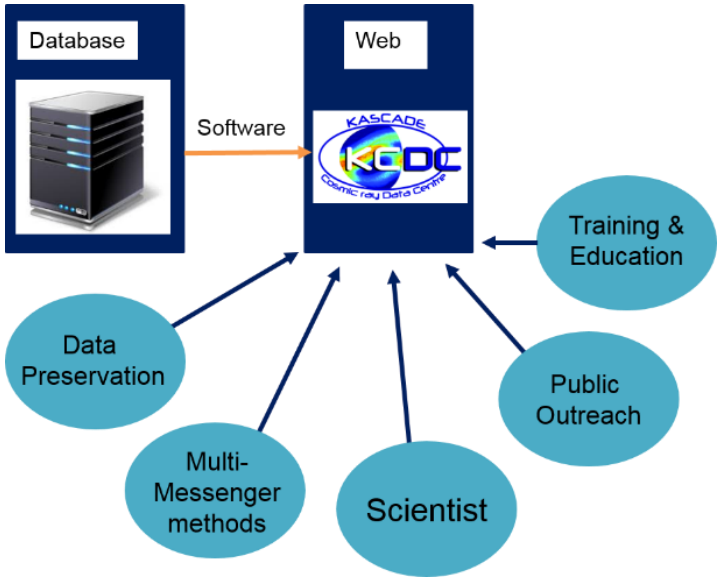
Motivation:

- 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)



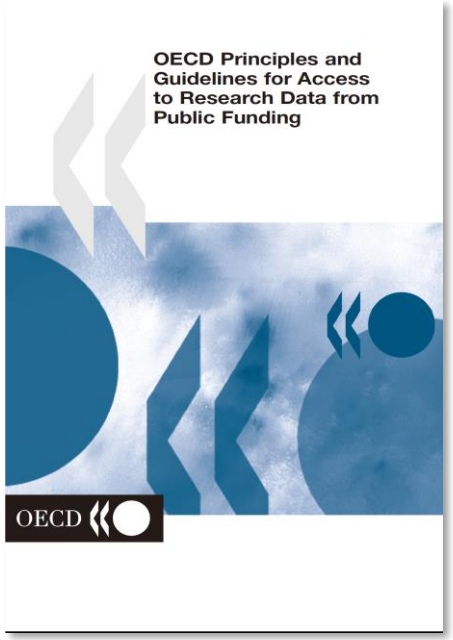
Important steps:

- 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

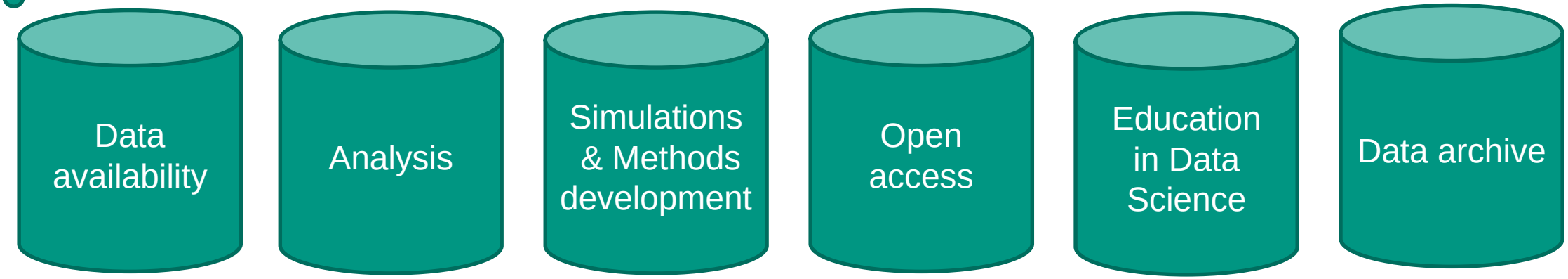
- Astroparticle Physics requests for multi-messenger analyses - this needs an **experiment-overarching** platform!



- **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:**

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).

- **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!

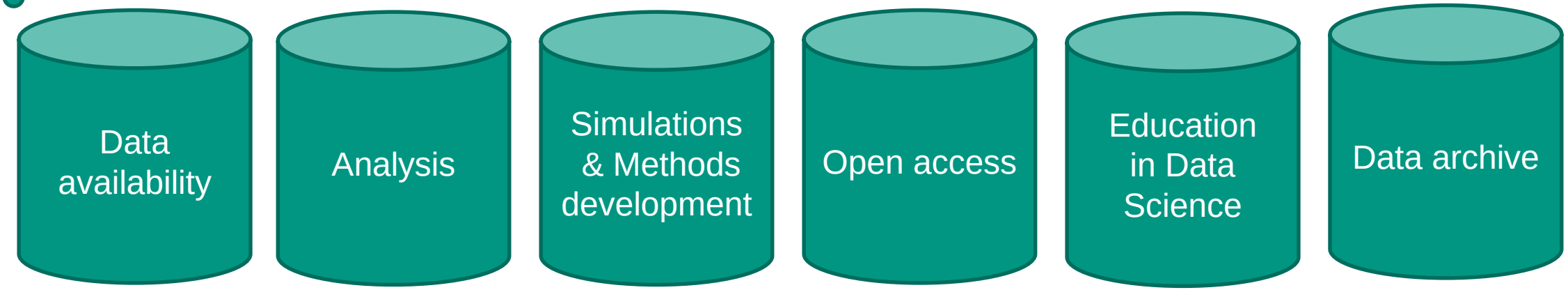
- **Education in data science:**

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).

Analysis and Data Centre in Astroparticle Physics

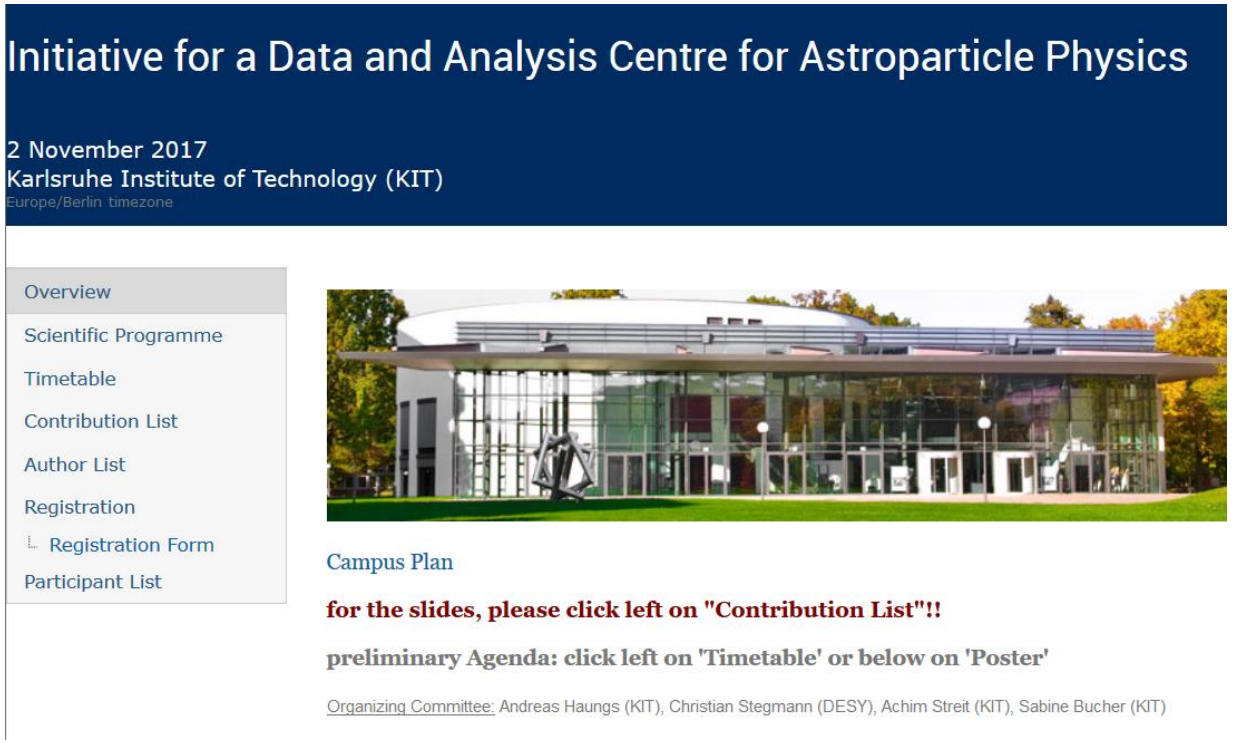


- **Data preservation** ----
  like DPHEP, KCDC
- **Metadata preservation** ----
  like KCDC
- **Data storage (archive)** ----
  like DPHEP, GridKa
- **Computing services (Grid vs. Cloud)** ---
  like CERN Tier-centres
- **Data access (policy, technology, rate)** ---
  like GridKa, KCDC
- **Training 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



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

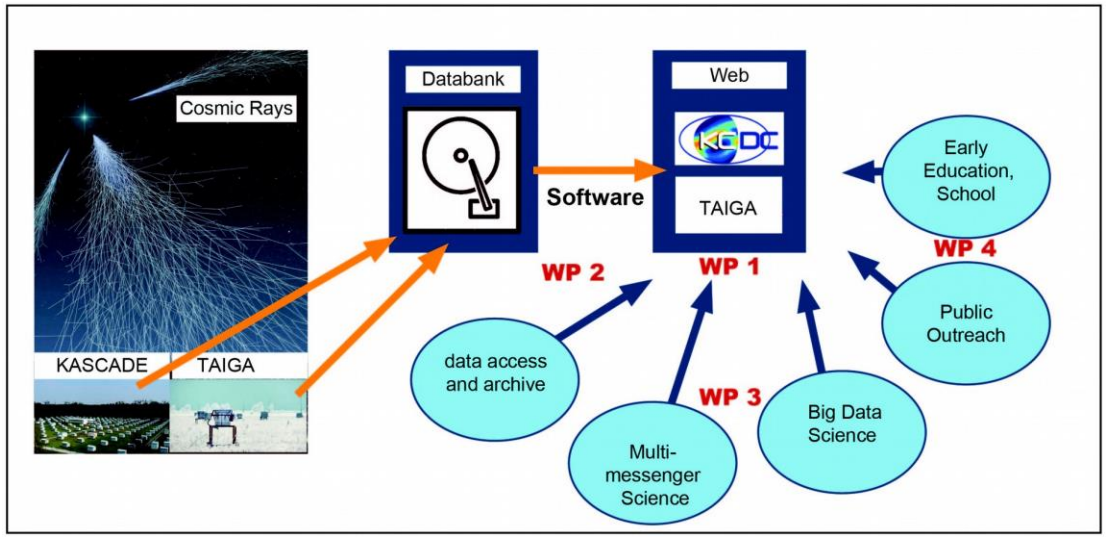
- Drafting white paper….

Karlsruhe-Russian Astroparticle Data Life Cycle Initiative

Open PhD positions!
Contact me…. ☺

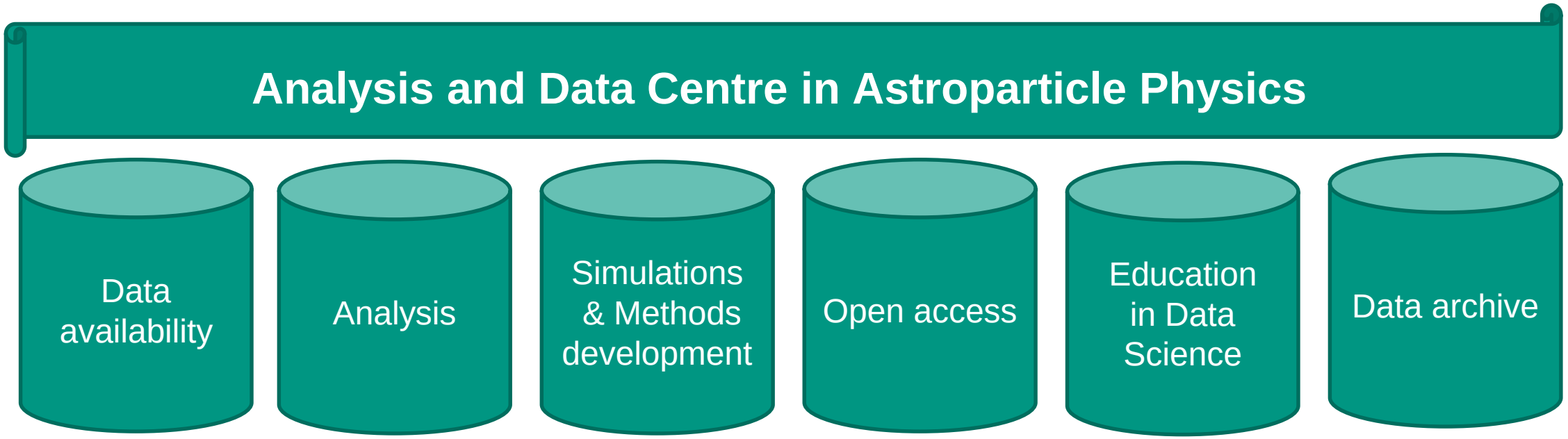
- **Basics**
  - project period (2018-2020)
  - Russia: SINP MSU, ISU, ISDCT SB RAS
    Germany: KIT, DESY
  - Team leaders: A. Kryukov (SINP MSU)
    and A. Haungs + A. Streit (KIT)

- **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



KRAD is a first step in extension and generalization of KCDC

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



Next steps:

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