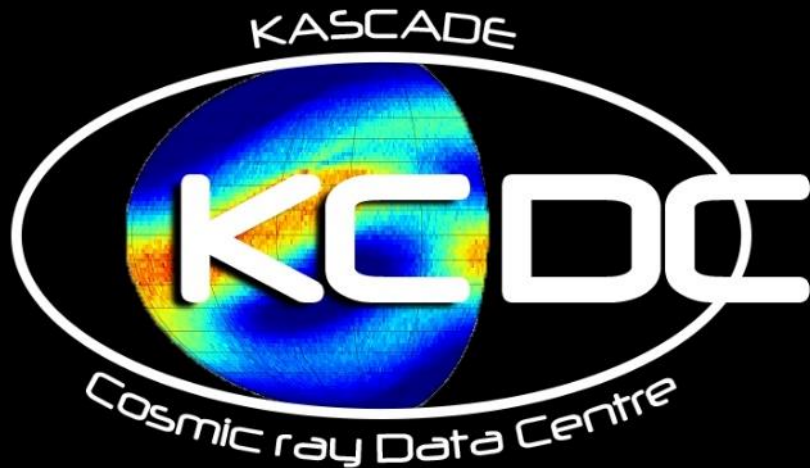


# KCDC:

## Data preservation and data publication in Astroparticle Physics

HAP workshop | Big Data Science in Astroparticle Physics  
Aachen, 20-22 February 2017

Andreas Haungs



Digitale Agenda der  
Bundesregierung

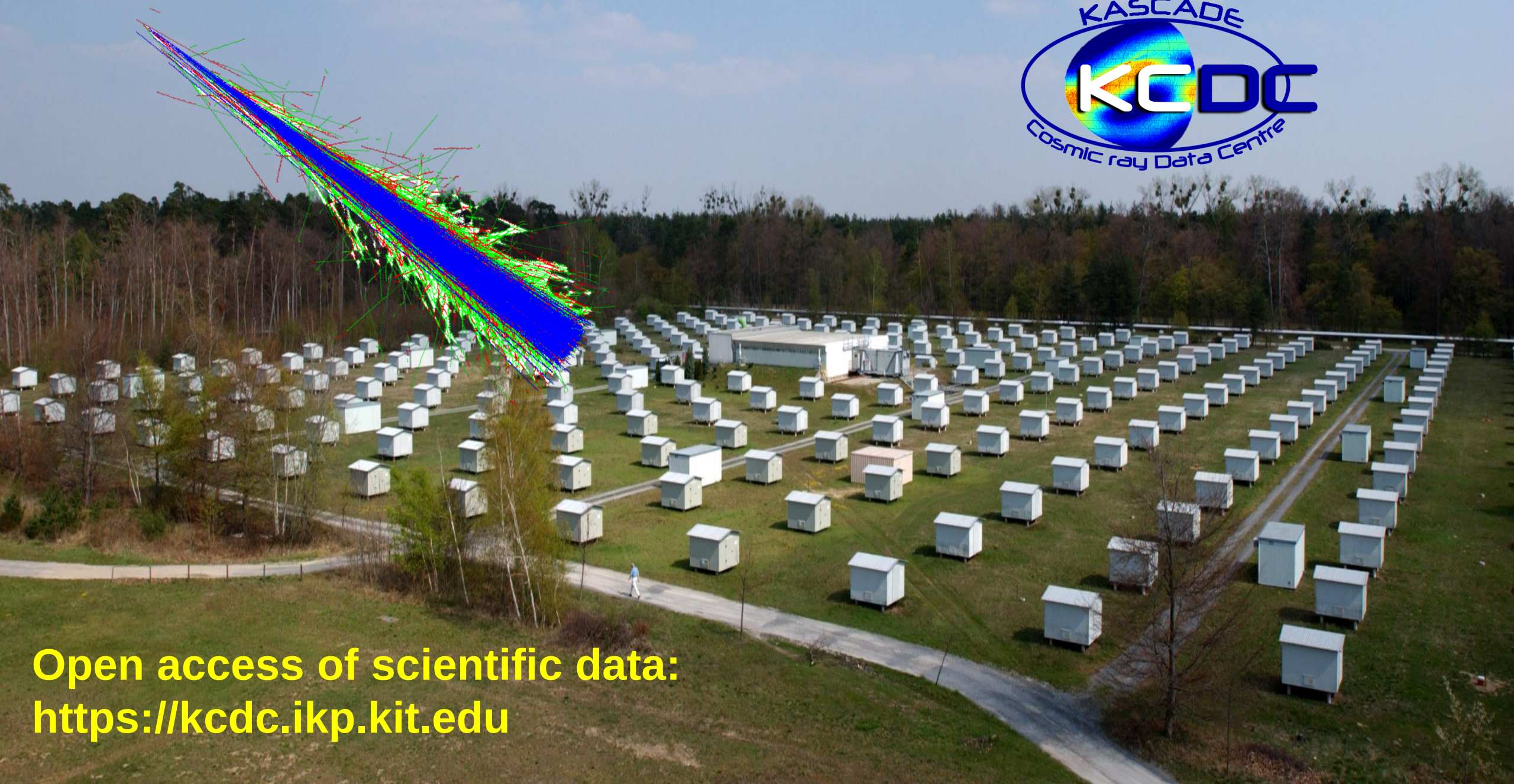


A stylized graphic with a blue and yellow color scheme. It features a yellow cable with multiple connectors leading towards a blue building with a chimney. The background is a mix of blue and yellow geometric shapes.

OECD Principles and  
Guidelines for Access  
to Research Data from  
Public Funding



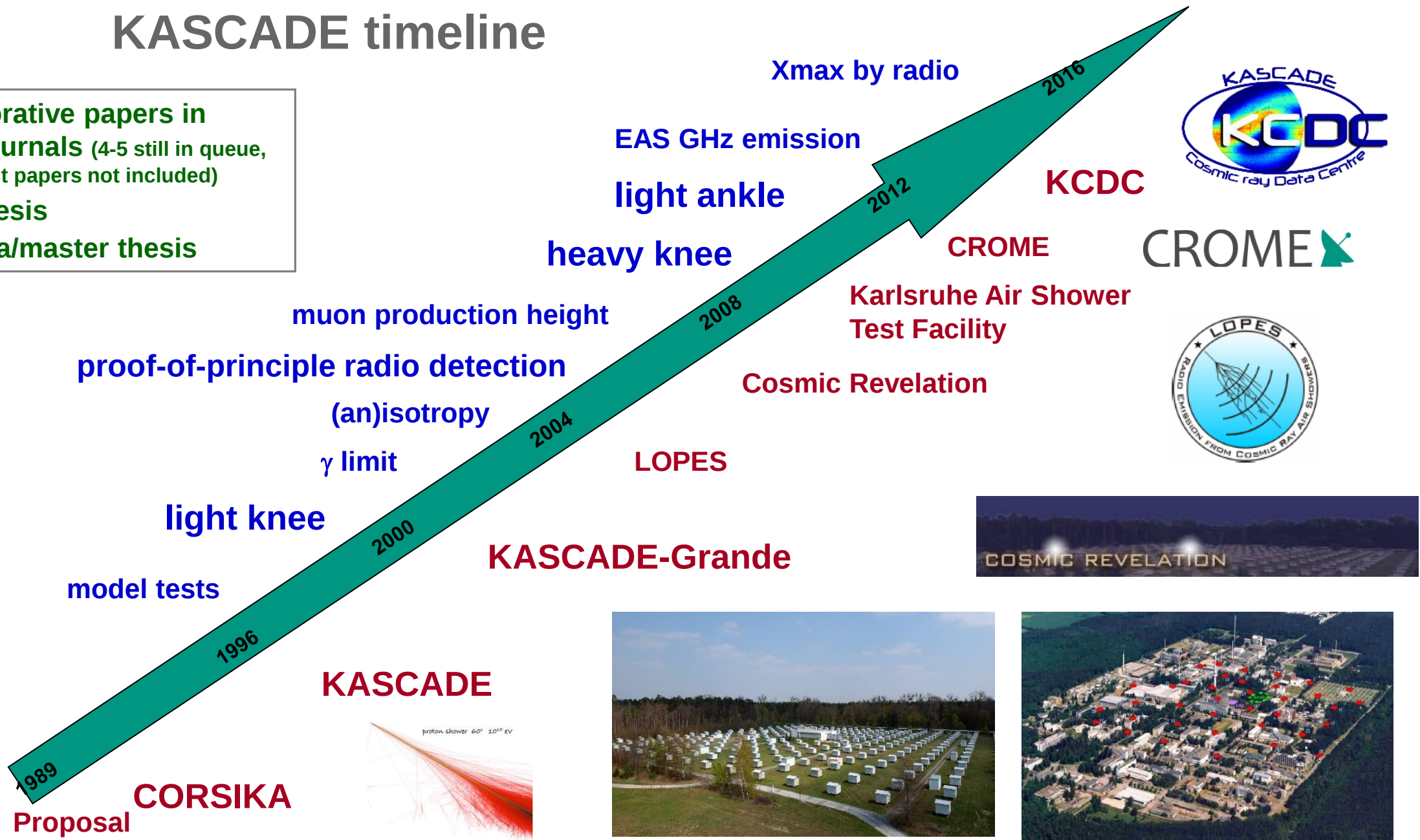
A graphic with a blue and white color scheme. It features a large blue circle with a white arrow pointing right. The background is a mix of blue and white geometric shapes.



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

# KASCADE timeline

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



# 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
- reliable data source
- guaranteed data quality

- **information platform**

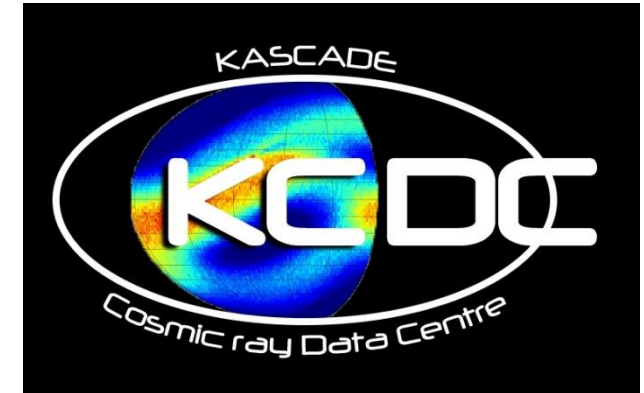
- experiment description
- meta information for data analysis
- physics background
- use of modern web technologies
- tutorials (focused on teachers and pupils)

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

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



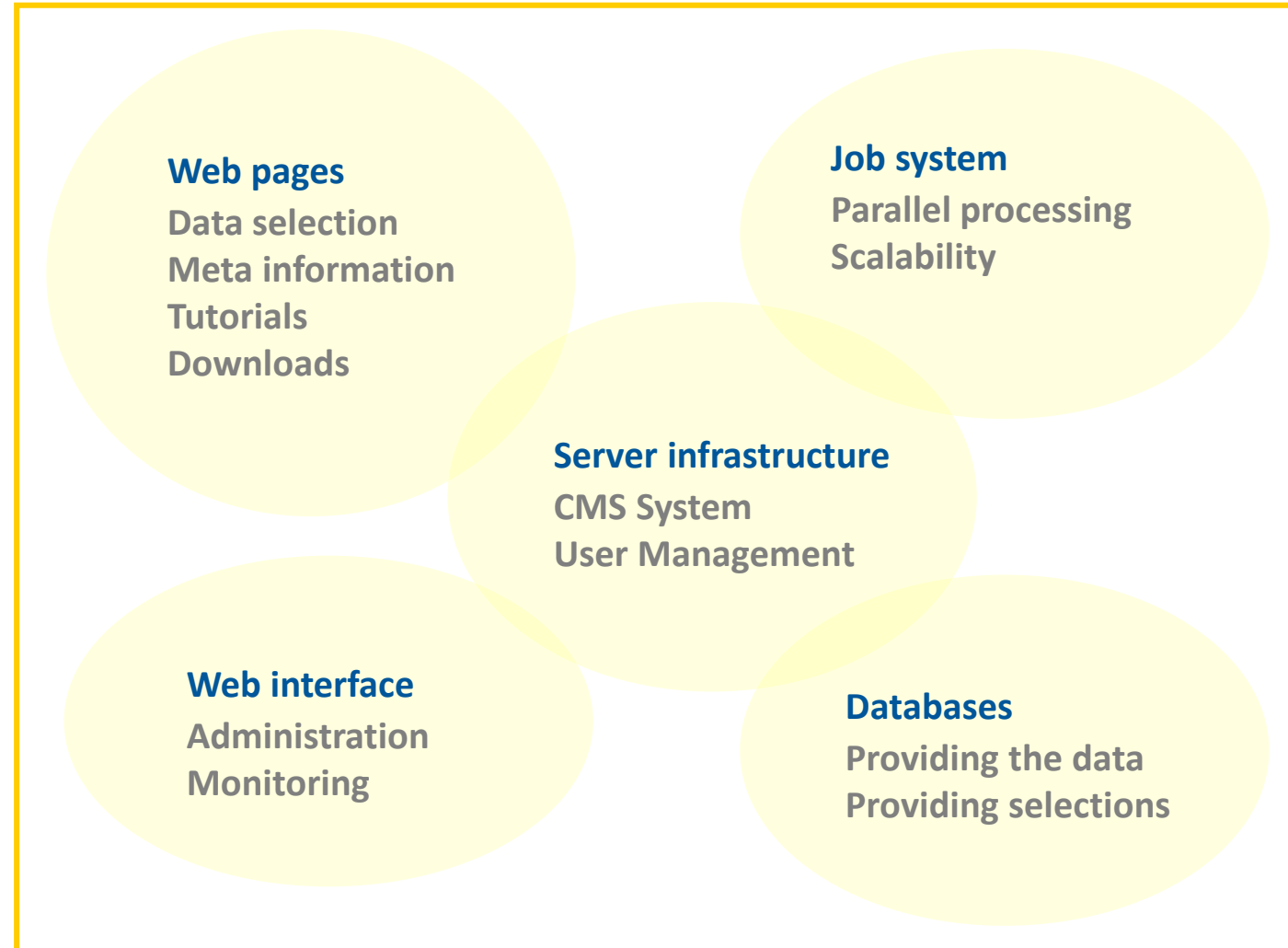
#KCDC\_KIT



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

# The Web Portal

- **open data publication**
  - follows the “Berlin Declaration on Open Data and Open Access”
  - explicitly requests the use of web technologies
- **free unlimited access for everyone**
  - scientific and non-scientific audience in focus, requires extensive documentation
- **modern technologies**
  - internet access & interactive data selections



# KCDC, the software

**providing a modern software solution  
for publishing KASCADE data  
for a general audience**

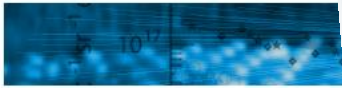
**In a second step: release the software  
as Open Source for free use by  
other experiments**

- Publication foreseen under Open Source License
- General software solution for open access to (astroparticle) data
- Following the concept of open access to research data
- Modular, flexible framework for data publication
- Good scalability (e.g. to large computing centers)
- Simple configuration via web interface
- Based solely on Open Source Software  
(Python, Django, HTML/Javascript and CSSdata provider)



# NABOO is released!

## 3.2.2017



KASCADE Cosmic Ray Data Centre (KCDC)



KCDC @KCDC\_KIT · 6s

Milestone release #KCDCversionNABOO is available providing all 433.000.000 events from KASCADE/KASCADE GRANDE! See [kcdc.ikp.kit.edu/announcements/...](http://kcdc.ikp.kit.edu/announcements/)



## Welcome to KCDC

The aim of the project **KCDC** (KASCADE Cosmic Ray Data Centre) is the installation and establishment of a public data centre for high-energy astroparticle physics based on the data of the KASCADE experiment. KASCADE was a very successful large detector array which recorded data during more than 20 years on site of the KIT-Campus North, Karlsruhe, Germany (formerly Forschungszentrum, Karlsruhe) at 49,1°N, 8,4°E; 110m a.s.l. KASCADE collected within its lifetime more than 1.7 billion events of which some 433.000.000 survived all quality cuts and are made available here for public usage.



KASCADE  
Karlsruhe Shower Core  
and Array Detector

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

KCDC OPEN BETA - VERSION NABOO.00 BA

Institute for Nuclear Physics  
(IKP)  
KIT Campus North

Address:

## ChangeLog - CLOSED BETA - Version : NABOO

### New & Modified Features

In the new release NABOO we have not only increased the amount of data published by a factor of more than 3 and added a new detector component (GRANDE), we have as well rebuilt the whole DataShop backend. The DataShop is now a plugin system based on a newly written software package which will also be made available for public usage.

### Details:

- we added about 290.000.000 events so now all 433.000.000 events recorded by the KASCADE/KASCADE-Grande experiment during 15 years of data taking are available for public usage.
- the newly added detector component is the **GRANDE Detector Array** (called GRANDE) with 9 new quantities.

These are:

- **Xc** - X-shower core position as reconstructed by GRANDE
  - **Yc** - Y-shower core position as reconstructed by GRANDE
  - **Ze** - zenith angle as reconstructed by GRANDE
  - **Az** - azimuth angle as reconstructed by GRANDE
  - **Nch** - number of charged particles in GRANDE
  - **Nmu** - number of muons from KASCADE as used in GRANDE
  - **Age** - shower age parameter as reconstructed by GRANDE
  - **GDeposit** - energy deposits of all charged particles in GRANDE/station
  - **GArrival** - arrival time of the first particle in GRANDE/station
- the former published quantities **e/γ-density** and **μ-density** have been replaced by **e/γ-deposit** and **μ-deposit** respectively which are directly measured and thus offer more flexible analysis
- the Preselections have been updated to the new data sets and extended
  - we updated to Ubuntu 16, mongoDB 3.2, Python 3.5, gunicorn 19, and RabbitMQ 3.6
  - we changed our storage concept in mongoDB to speed up the processing of the requests
  - the KCDC-Manual has been extended
  - we cleared inactive accounts and user histories according privacy protection guidelines

# KCDC data shop

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

**KIT**  
Karlsruhe Institute of Technology

KASCADE Cosmic Ray Data Centre (KCDC) / Open B

**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 <input type="button" value="Add Cut"/>         |
|                      | GRANDE              | <input type="checkbox"/> Air Pressure range: 980 to 1040 hPa <input type="button" value="Add Cut"/>         |
|                      |                     | <input type="checkbox"/> DateTime range: 1998-05-08 to 2013-01-15 <input type="button" value="Add Cut"/>    |
|                      |                     | <input type="checkbox"/> Global Time range: 8.945e+8 to 1.255e+9 sec <input type="button" value="Add Cut"/> |
|                      |                     | <input type="checkbox"/> Mt range: 0 to 9.99e+8 m <input type="button" value="Add Cut"/>                    |
|                      |                     | <input checked="" type="checkbox"/> Run Number range: 877 to 7417 <input type="button" value="Add Cut"/>    |
|                      |                     | <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

## 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

The screenshot shows the KCDC website interface. At the top, there is a navigation bar with links: [ haungs ] | HOME | KIT | IKP | IMPRESSUM | ADMIN | LOGOUT. The main header features the KIT logo (Karlsruhe Institute of Technology) on the left and the KASCADE Cosmic Ray Data Centre (KCDC) logo on the right. Below the header, a banner image displays a cosmic ray shower with the text 'KASCADE Cosmic Ray Data Centre (KCDC) / Open  $\beta$ '. The left sidebar contains a menu with the following items: KCDC Homepage, KCDC Motivation, Regulations (highlighted), Information, Announcements, FAQs, and DATA Shop. The main content area is titled 'Regulations - Legal Aspects of KCDC' and contains the following text:

**End User Licence Agreement for using the KCDC webportal and the KCDC data (EULA)**

This EULA provides the rights and duties of the usage of the KASCADE Cosmic Ray Data Centre (KCDC) webportal (hereinafter called WEBPORTAL) as well as the corresponding scientific KCDC data (hereinafter called DATA). The Karlsruhe Institute of Technology (KIT) is the owner of the WEBPORTAL which contains DATA, as well as printable materials about KCDC and online or electronic documentation about KCDC (hereinafter called DOCUMENTS), and related modules (hereinafter called SERVICES) of KCDC.

Please read this EULA carefully. By using the WEBPORTAL or by using any SERVICES or by downloading DATA, You (You, as the licensee, are hereinafter called YOU) agree that this EULA is enforceable like any written contract signed by YOU. If YOU do not agree to all of the terms of this EULA, click on the button that indicates that YOU do not agree to accept the terms of this EULA (if applicable) and do not continue the use of the WEBPORTAL, the provided DATA, or the

# 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  $\beta$

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.

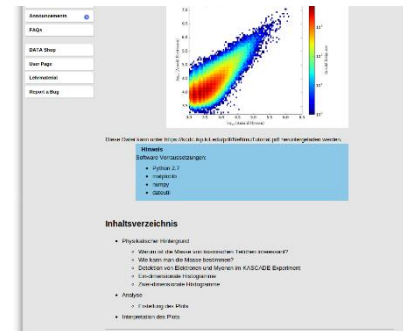
here you get: [exercise](#) - [tutorial](#) - [solution](#)

**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.

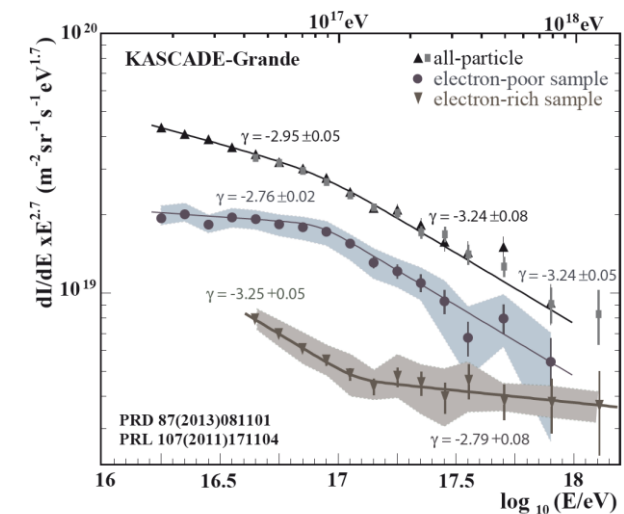
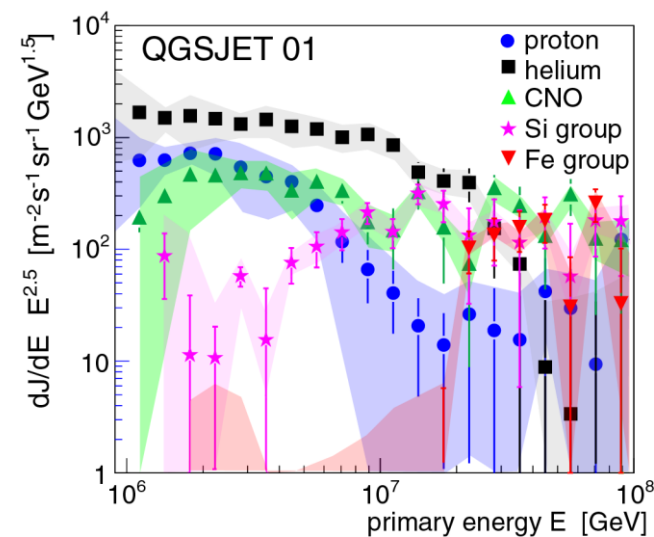
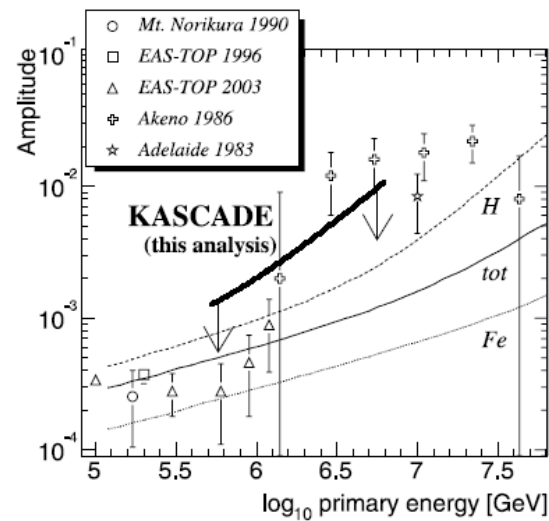
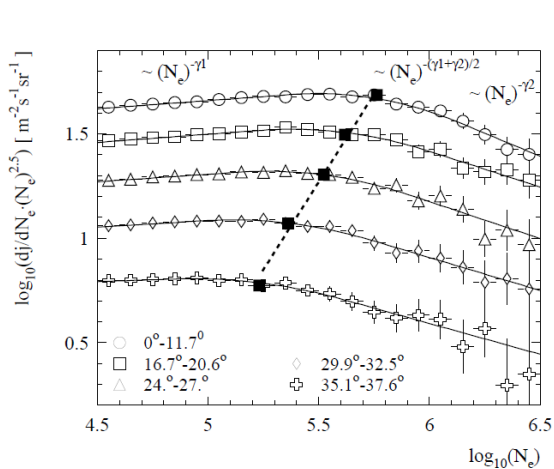
here you get: [exercise](#) - [tutorial](#) - [solution](#)

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



This is the data set for analysis works like

- **Astroparticle Physics 19 (2003) 703-714**  
*Measurement of Attenuation and Absorption Lengths with the KASCADE Experiment*
- **The Astrophysical Journal 608 (2004) 865-871**  
*Search for Cosmic-Ray Point Sources with KASCADE*
- **Astroparticle Physics 24 (2005) 1-25**  
*KASCADE Measurements of energy spectra for elemental groups of cosmic rays: Results and open problems*
- **Physical Review Letters 107 (2011) 171104**  
*Kneelike Structure in the Spectrum of the Heavy Component of Cosmic Rays Observed with KASCADE-Grande*
- etc.



# KCDC in context of Big Data Science in Astroparticle Physics

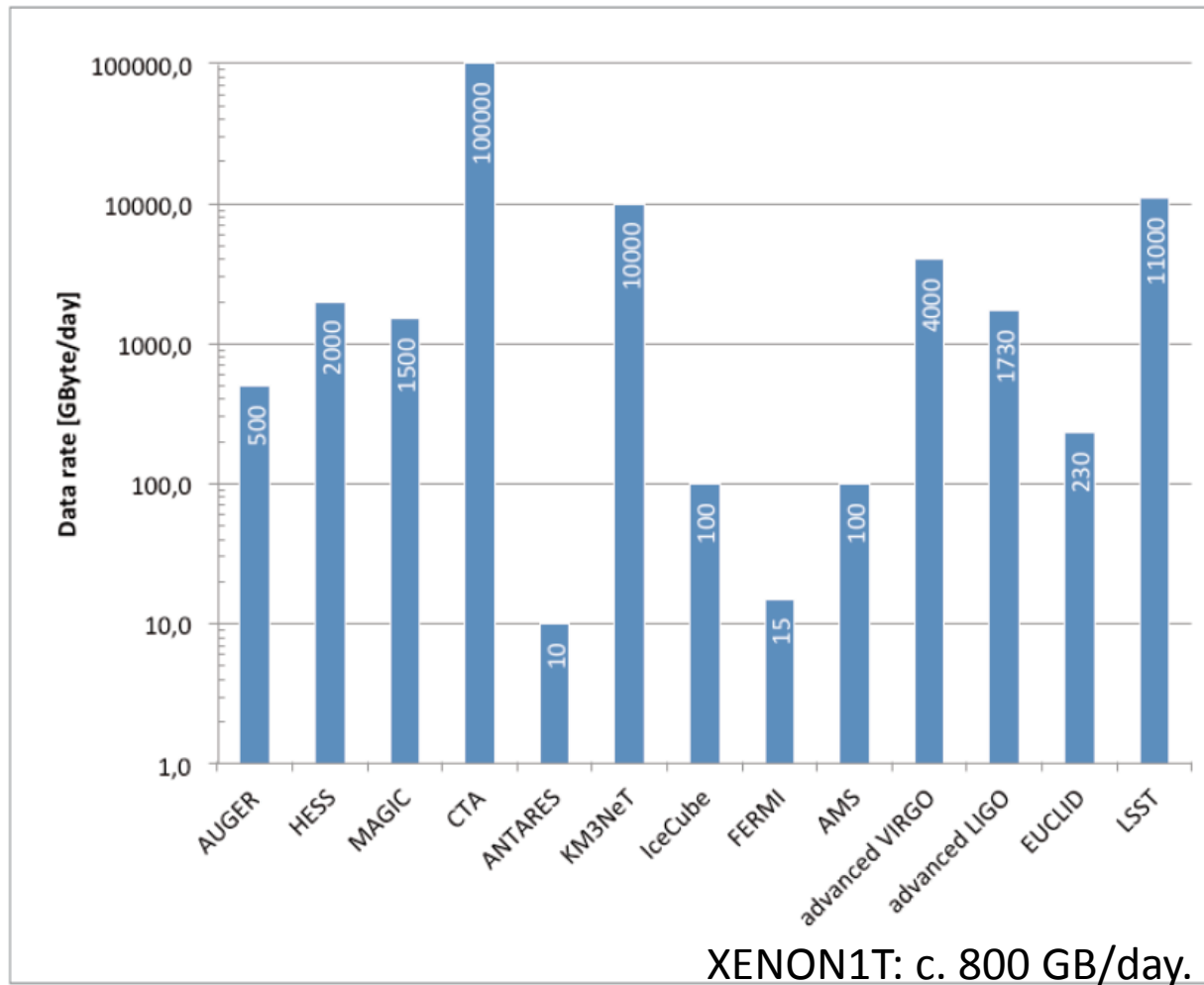
The Astroparticle Physics community needs to continue discussion on its

- Data preservation
- Metadata preservation
- Data storage (archive)
- Computing services (Tier-centres for astroparticle physics)
- Data access (policy, technology, rate)
- Training on Data use (maintenance, tutorials)
- Data analysis, Simulation, modeling
- Data science (tools, e.g. deep learning)
- Data publication / Outreach
- Data exchange
- Data catalogues
- .....

KCDC can serve as concept  
for some of these issues but  
not for all!

....in order to provide a coherent and efficient way (strategy) in using Big Data

# Computing in Astroparticle Physics (Astro-Grid / Astro-Cloud)



Source: APPEC brochure on Computing, 2016

→ Do we need an own Astroparticle Physics computing infrastructure?

- independent of particle physics?
- Grid or Cloud or other technology?
- Use of commercial provider (amazon, google, ...)?

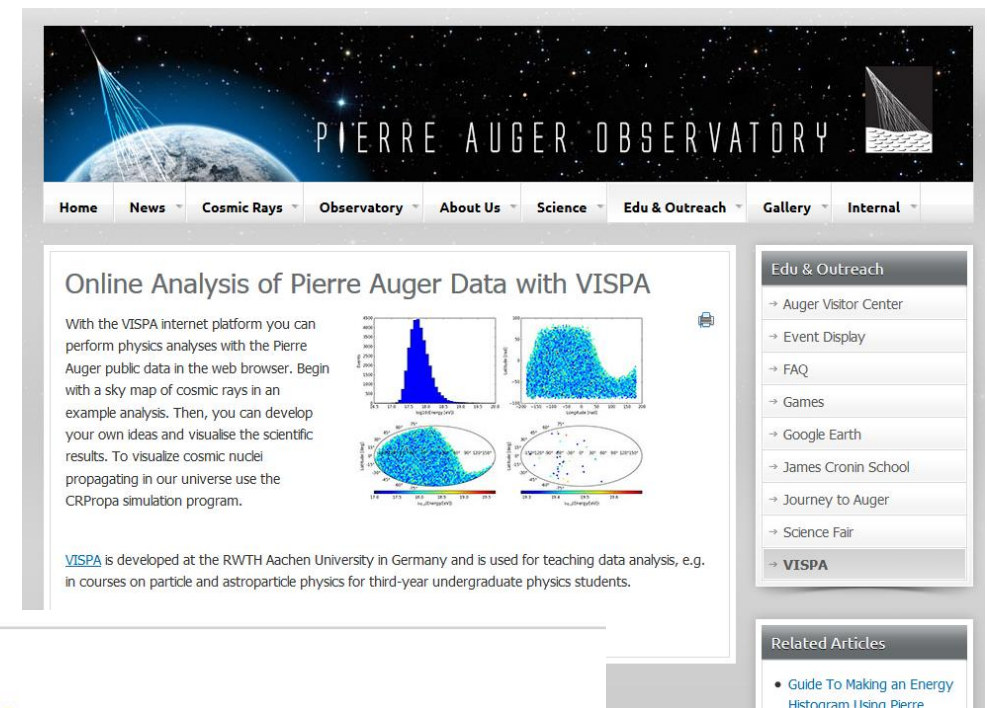
# Outreach

- **VISPA** to analyze Auger data  
Learning Deep Learning 😊  
Algorithms & data analysis in own browser  
Example analysis  
Writing own algorithms  
Visualizing own results

- **Masterclasses**  
Netzwerk Teilchenwelt  
(Auger & IceCube)

- **Cosmic Days**

- ....



The screenshot shows the Pierre Auger Observatory website. The header features the observatory's name and a navigation menu with links: Home, News, Cosmic Rays, Observatory, About Us, Science, Edu & Outreach, Gallery, and Internal. The main content area is titled "Online Analysis of Pierre Auger Data with VISPA" and includes a description of the platform's capabilities, accompanied by four plots showing cosmic ray data analysis results. A sidebar on the right lists "Edu & Outreach" resources like the Auger Visitor Center, Event Display, FAQ, Games, Google Earth, James Cronin School, Journey to Auger, Science Fair, and VISPA. Below this, a "Related Articles" section mentions a guide to making an energy histogram using Pierre.



The screenshot shows the Netzwerk Teilchenwelt website. The header features the logo and the text "NETZWERK TEILCHENWELT ..... QUARKS, ELEKTRONEN & CO...". Below the header is a navigation menu with links: DAS PROJEKT, AKTUELLES, MITMACHEN, ANGEBOTE, STANDORTE, MATERIAL, and FORUM. The main content area is titled "Astroteilchenphysik" and includes a large image of a cosmic structure. To the right of the image is a sidebar with links: "Dow", "Vide", "Teil", "Link", "Inter", "Wel", "Astr", "Karr", "Zeit", "Cos", and "Artik". Below the image, the text reads: "Astroteilchenphysik und Elementarteilchenphysik folgen einem gemeinsamen Ziel: die Struktur und Entstehung der Materie zu erklären. Kosmische Teilchen spielen dabei eine wichtige Rolle:"

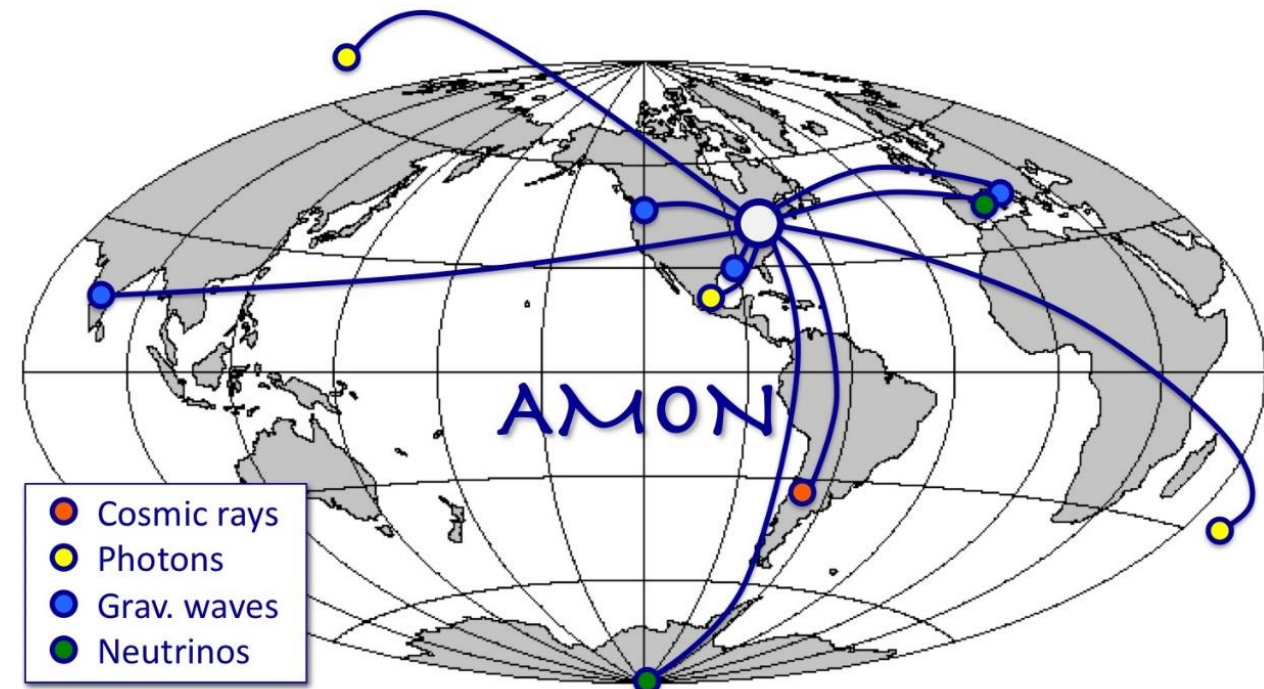
# Exchange of Data / Alert systems

<http://amon.gravity.psu.edu/>

## Members and Prospective Members

| Observatory                                      | Contact             | Letter of Collaboration | MoU in Review | MoU Signed |
|--------------------------------------------------|---------------------|-------------------------|---------------|------------|
| ANTARES                                          | Juergen Brunner     | ✓                       | ✓             | ✓<br>MOU   |
| Auger                                            | Miguel Mostafa      | ✓                       | ✓             | ✓<br>MOU   |
| FACT                                             | Adrian Biland       |                         |               | ✓<br>MOU   |
| Fermi                                            | Julie McEnery       | ✓                       |               |            |
| HAWC                                             | Ignacio Taboada     | ✓                       | ✓             | ✓<br>MOU   |
| IceCube                                          | Doug Cowen          | ✓                       | ✓             | ✓<br>MOU   |
| Las Cumbres Observatory Global Telescope (LCOGT) | Todd Boroson        | ✓                       | ✓             | ✓<br>MOU   |
| LIGO                                             | Gabriela Gonzalez   | ✓                       |               |            |
| Large Millimeter Telescope                       | Alberto Carramiñana | ✓                       | ✓             | ✓          |
| MASTER                                           | Vladimir Lipunov    |                         |               | ✓<br>MOU   |
| Palomar Transient Factory                        | Tom Prince          | ✓                       |               |            |
| Swift                                            | Scott Barthelmy     | ✓                       | ✓             | ✓          |
| VERITAS                                          | Abe Falcone         | ✓                       | ✓             | ✓          |

Membership to AMON is open to any relevant facility, subject to signing of the [AMON MOU](#).



# Data Catalogues

<http://www.re3data.org/>

- Sample and links to repositories of scientific data, mostly results, not the “data”.

e.g., search for “Cosmic Rays”:

Found 7 result(s):

1. [World Data Center for Cosmic Rays](#) [WDCCR](#)
2. [KASCADE Cosmic Ray Data Centre](#) [KCDC](#)
3. [Spitzer Science Archive](#) [SHA](#)
4. [World Data Center for Solar-Terrestrial Physics, Moscow](#)
5. [Virtual Space Science Observatory](#) [VSSO](#)
6. [LAADS Web](#)
7. [High Energy Astrophysics Science Archive Research Center](#)

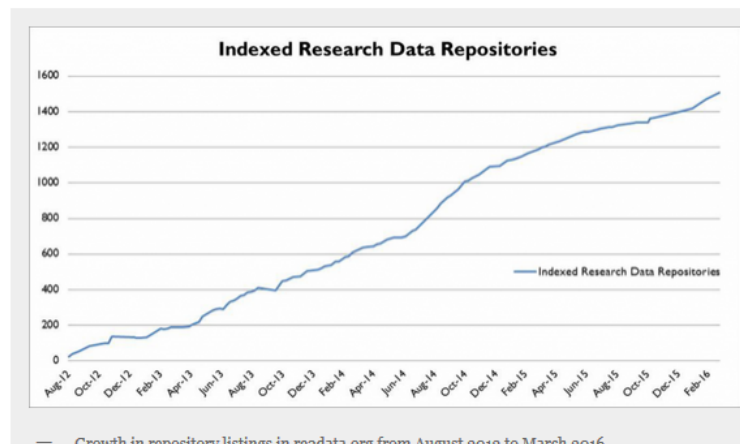
**re3data.org**  
REGISTRY OF RESEARCH DATA REPOSITORIES

Home Search Browse Suggest FAQ About Schema API Contact Legal notice / Impressum

## re3data.org Reaches a Milestone & Begins Offering Badges

Posted on April 13, 2016 by [re3data.org team](#)

re3data.org has reached a milestone of identifying and listing 1,500 research data repositories, making it the largest and most comprehensive registry of data repositories available on the web. It has grown steadily since its launch four years ago to cover a wide range of disciplines from around the world.



SEARCH

A SERVICE BY



PARTNERS



# Example Particle Physics: Data Preservation



| Preservation Model |                                                                                     | Use Case                                                       |  |
|--------------------|-------------------------------------------------------------------------------------|----------------------------------------------------------------|-------------------------------------------------------------------------------------|
| 1                  | Provide additional documentation                                                    | Publication related info search                                | Documentation                                                                       |
| 2                  | Preserve the data in a simplified format                                            | Outreach, simple training analyses                             | Outreach                                                                            |
| 3                  | Preserve the analysis level software and data format                                | Full scientific analysis, based on the existing reconstruction | Technical Preservation Projects                                                     |
| 4                  | Preserve the reconstruction and simulation software as well as the basic level data | Retain the full potential of the experimental data             |                                                                                     |

## Data Preservation:

- Define objectives of the data persistency in HEP.
- Exchange information concerning the analysis models: abstraction, software, documentation etc. and identify coherence points.
- Hardware and software persistency status.
- Review possible funding programs and other related international initiatives.
- Converge to a common set of specifications in a document that will constitute the basis for future collaborations.

<https://www.dphep.org/>

D. South, DPHEP collaboration

# Example Astronomy: Strasbourg astronomical Data Center



**Centre de Données astronomiques de Strasbourg**  
*Strasbourg astronomical Data Center*

<http://cds.u-strasbg.fr/>



Entry point to all services 



Object database 



Catalogue database 



Interactive sky atlas 

Combines many of the earlier mentioned issues:

- User Portal
- Data bases
- Tools
- Catalogues...

→ What is different in astroparticle physics?

Diversity of Data, calibration, format, analysis, ...

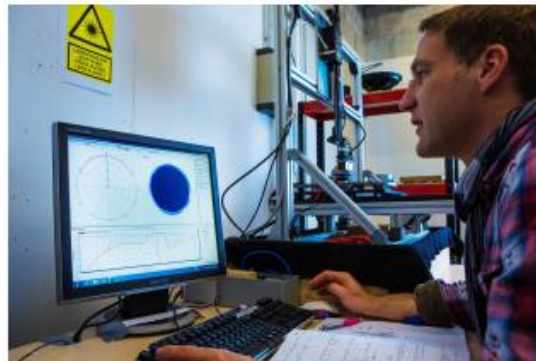
![also very different for low energy astroparticle physics experiments]!

## Handlungsfelder

1. Digitale Infrastrukturen
2. Digitale Wirtschaft und digitales Arbeiten
3. Innovativer Staat
4. Digitale Lebenswelten in der Gesellschaft gestalten
5. Bildung, Forschung, Wissenschaft, Kultur und Medien
6. Sicherheit, Schutz und Vertrauen für Gesellschaft und Wirtschaft
7. Europäische und internationale Dimension der Digitalen Agenda

## Zugang zu Wissen als Grundlage für Innovation sichern

Wir werden die Rahmenbedingungen für einen ungehinderten Informationsfluss, insbesondere in der Wissenschaft, verbessern. Diese dienen dazu, die Potenziale für Wissenschaft, Forschung und Bildung voll zu nutzen. ➤ mehr



**Die #Digitale Agenda ist Auftakt für einen offenen Prozess, den alle gesellschaftlichen Gruppen aktiv mitgestalten sollen.**



## Chapters

....

### **V. Bildung, Forschung, Wissenschaft, Kultur und Medien**

1. Digitalen Wandel in der Wissenschaft forcieren
2. Zugang zu Wissen als Grundlage für Innovation sichern

....

# Data Center in Astroparticle Physics

Data  
availability

Analysis

Simulations  
& Methods  
development

open access

Data archive

- Data preservation ---- like DPHEP, KCDC
- Metadata preservation ---- like KCDC
- Data storage (archive) ---- like DPHEP
- 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 exchange --- like AMON, CREDO?
- Data catalogues --- like Re3Data

▪ .....

# Initiative for a (global) Data Center in Astroparticle Physics ?

- First for High-Energy Astroparticle Physics?
  - Larger facilities, more data, request for multi-messenger analysis?
- **Tasks**
  - Provide sustainable access to scientific data
  - Archiving of Data and Meta-Data
  - Providing analysis tools
  - Development area for multi-messenger analyses (e.g. Deep Learning)
  - Platform for communication and exchange within Astroparticle Physics
- **Elements**
  - Advancement, generalization of KCDC
  - In direction of a virtual Observatory (like in astronomy)
  - In direction of Grid and DPHEP (like in particle physics)
  - „Digitale Agenda der Bundesregierung“
  - OECD Principles and Guidelines for Access to Research Data from Public Funding
- High demand in (German and international) Community ! (?)
- Needs dedicated efforts and resources, i.e. concerted action

