Karlsruhe Institute of Technology

HELMHOLTZ GEMEINSCHAFT

Allianz für Astroteilchenphysik

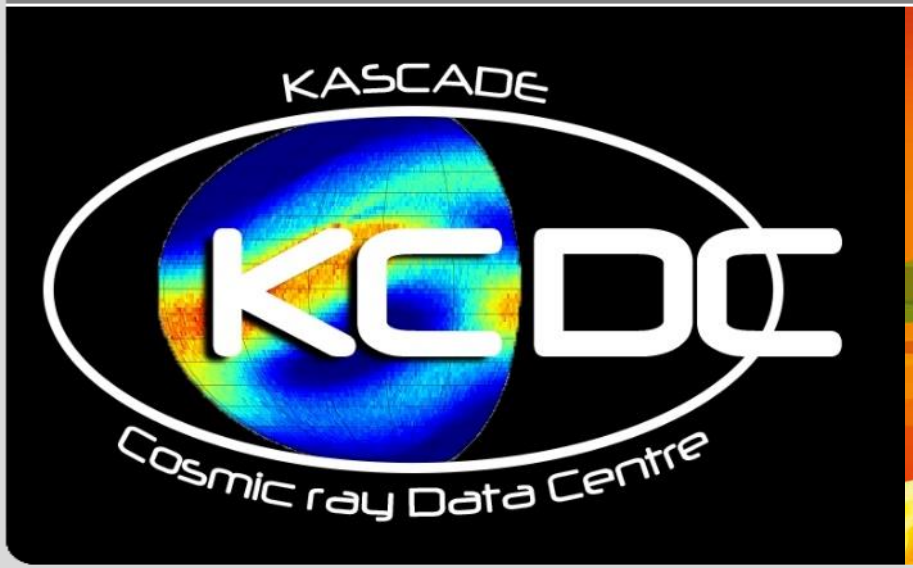
KAT. Komitee für Astro.Teilchen.Physik

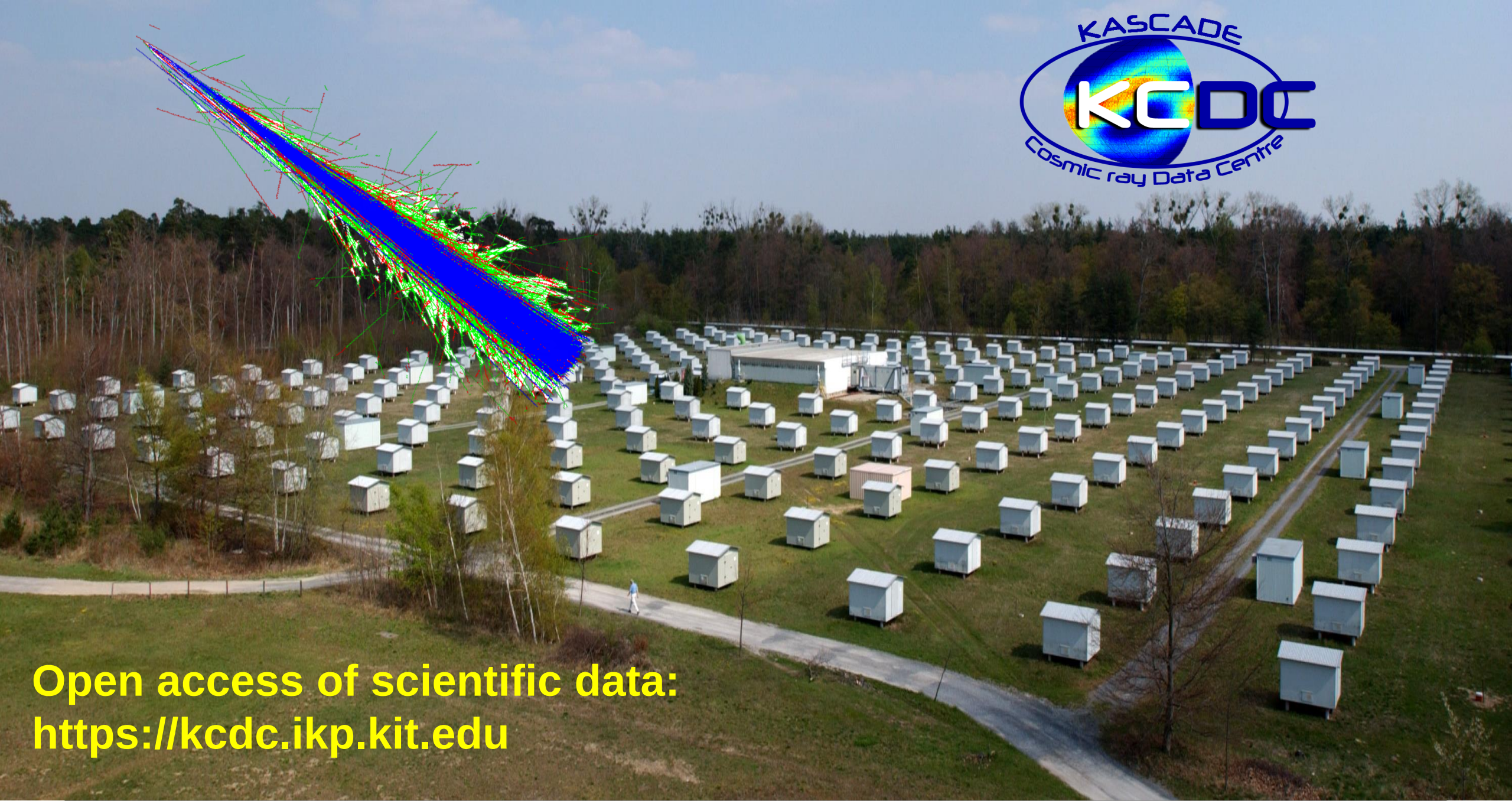
KCDC:

Data preservation and data publication in Astroparticle Physics

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

Andreas Haungs




KASCADE
KCDC
Cosmic ray Data Centre
Open access of scientific data:
https://kcdc.ikp.kit.edu

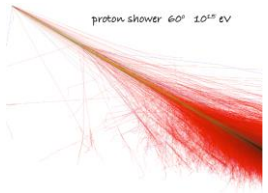
KASCADE timeline

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

1989 Proposal

1996 CORSIKA

2000 KASCADE



2004 model tests

light knee

KASCADE-Grande

2008 γ limit

(an)isotropy

proof-of-principle radio detection

muon production height

LOPES

2012 heavy knee

light ankle

EAS GHz emission

Cosmic Revelation

Karlsruhe Air Shower Test Facility

CROME

2016 Xmax by radio

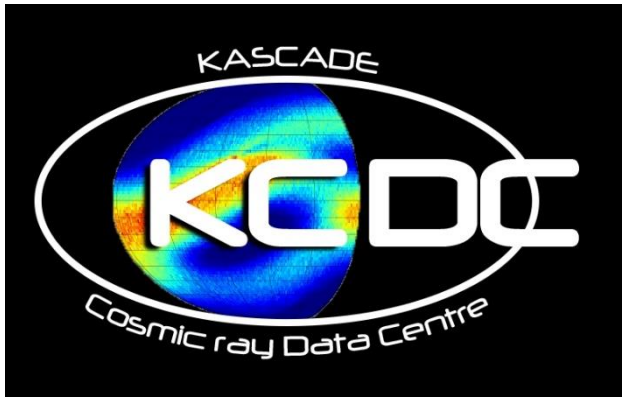
KCDC

KCDC in a nutshell

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

#KCDC_KIT

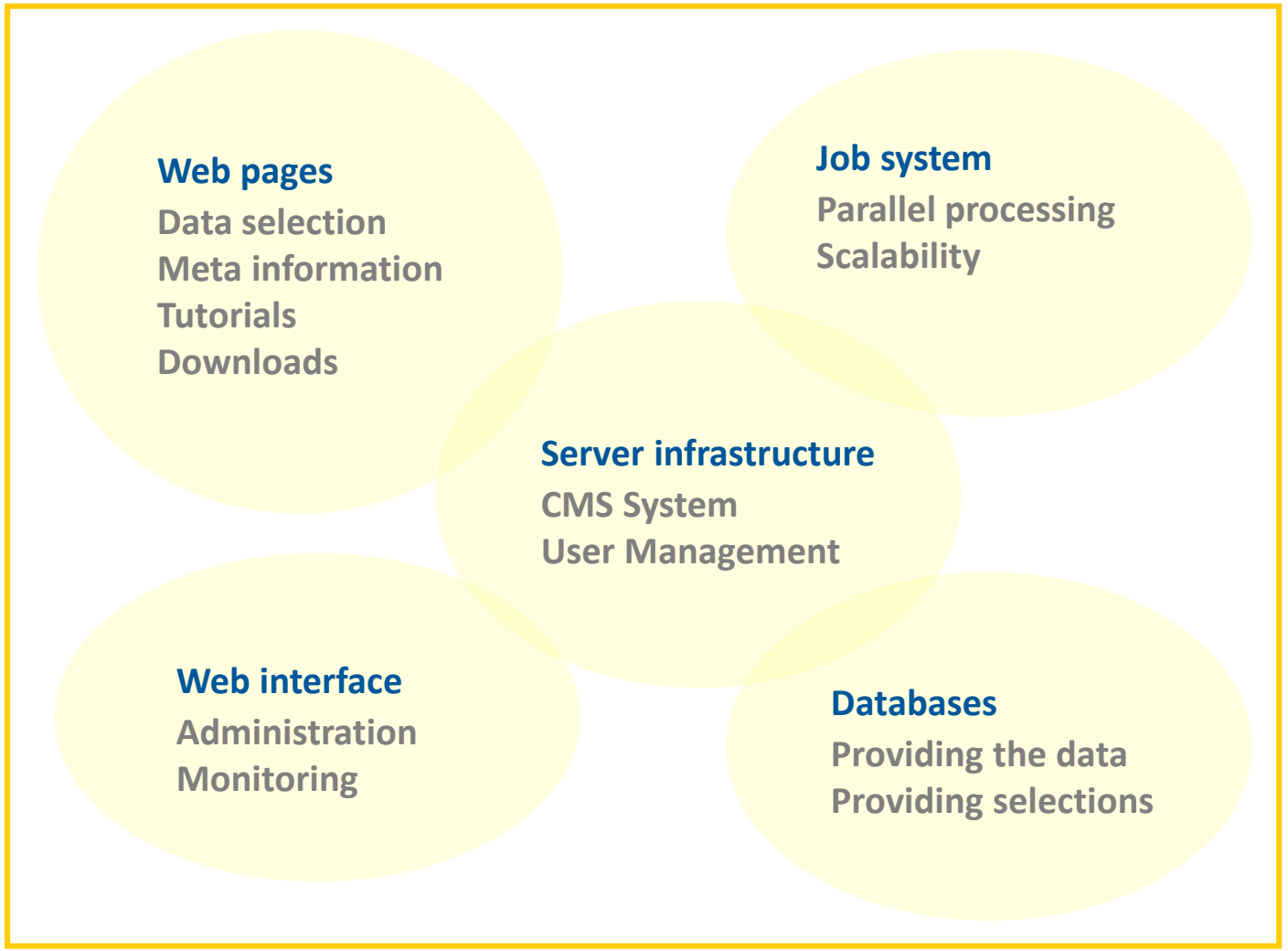
- **data provider**
  - follows the “Berlin Declaration on Open Data and Open Access”
  - free, unlimited, open access to KASCADE cosmic ray data
  - selection of fully calibrated quantities
  - reliable data source
  - guaranteed data quality
- **information platform**
  - experiment description
  -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



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

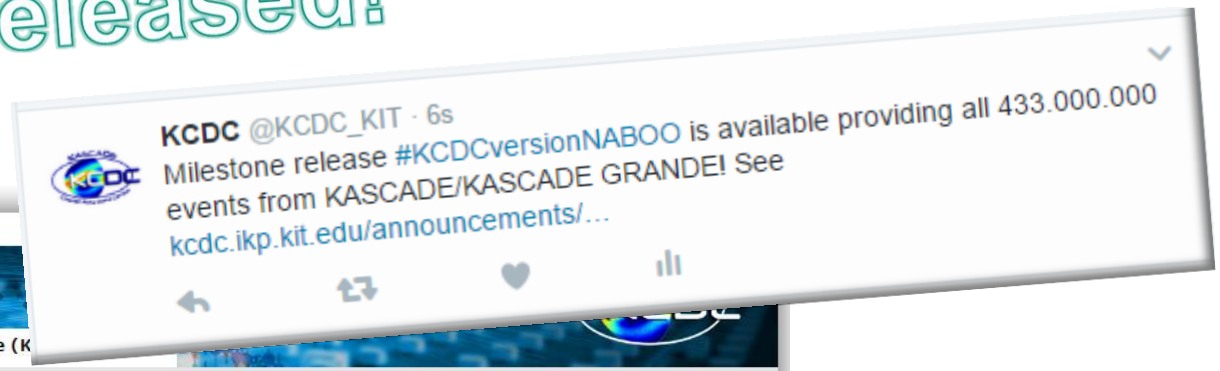


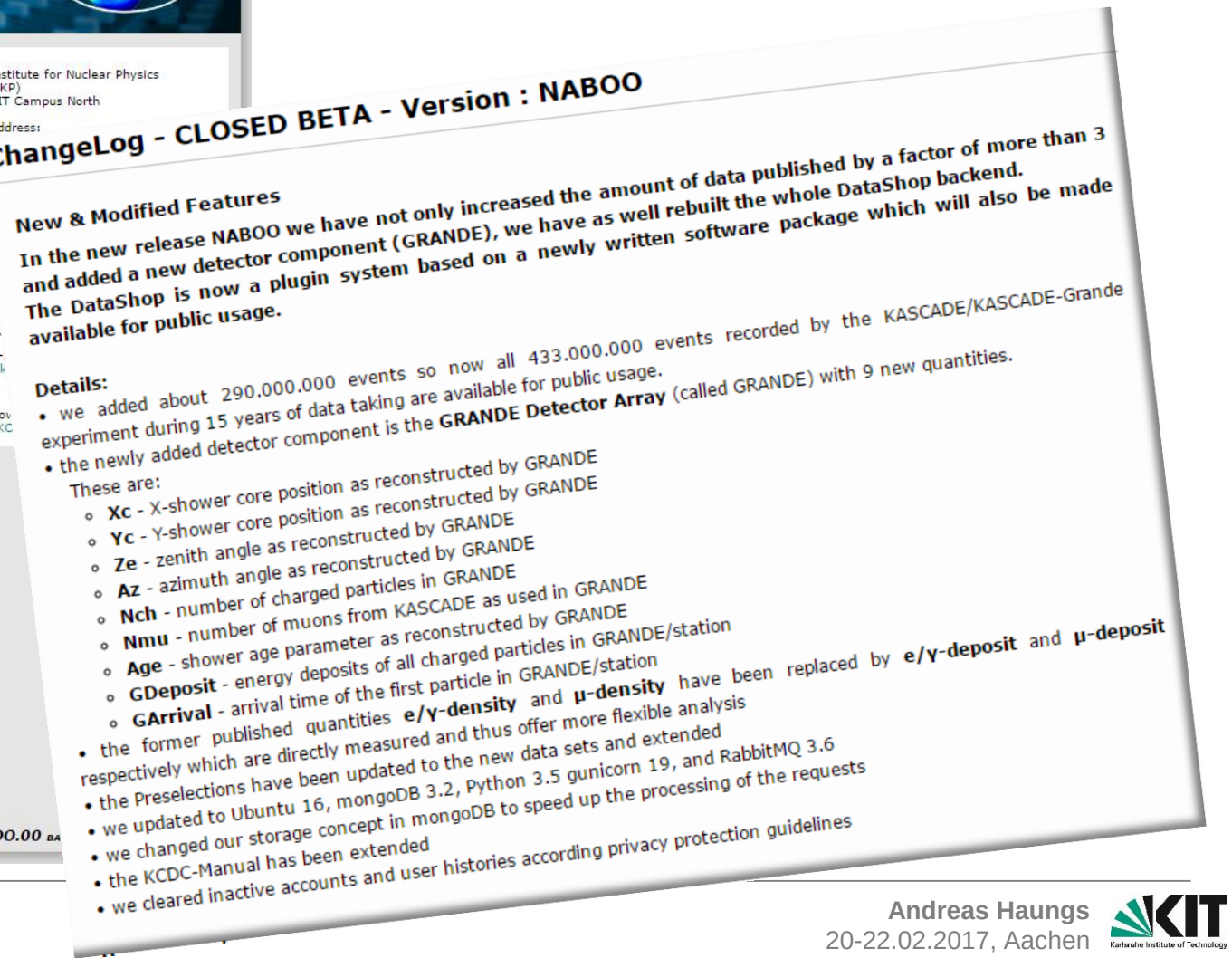
KCDC, the software

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

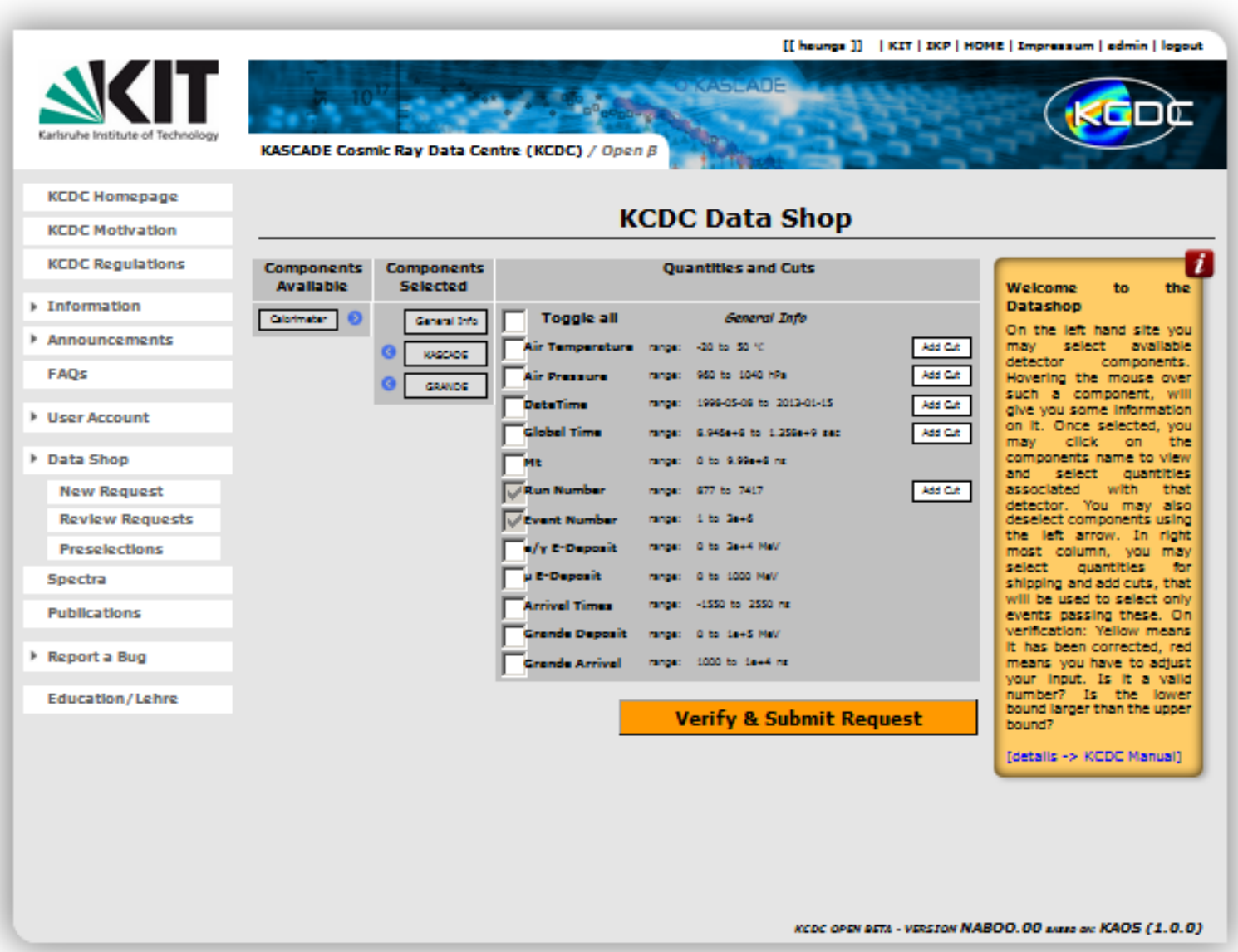
- Publication foreseen under Open Source License
- General software solution for open access to (astroparticle) data
- Following the concept of open access to research data
- Modular, flexible framework for data publication
- Good scalability (e.g. to large computing centers)
- Simple configuration via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

NABOO is released!
3.2.2017





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

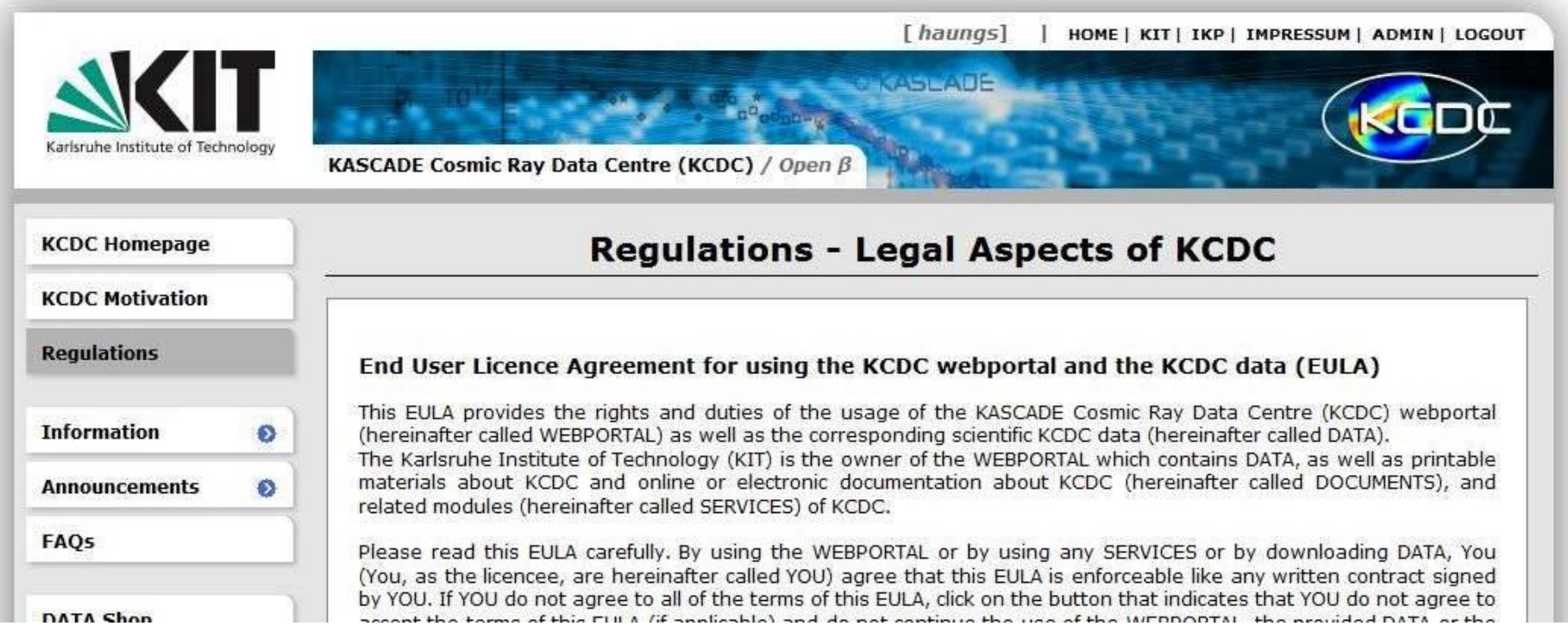
Law and Order

open data publication

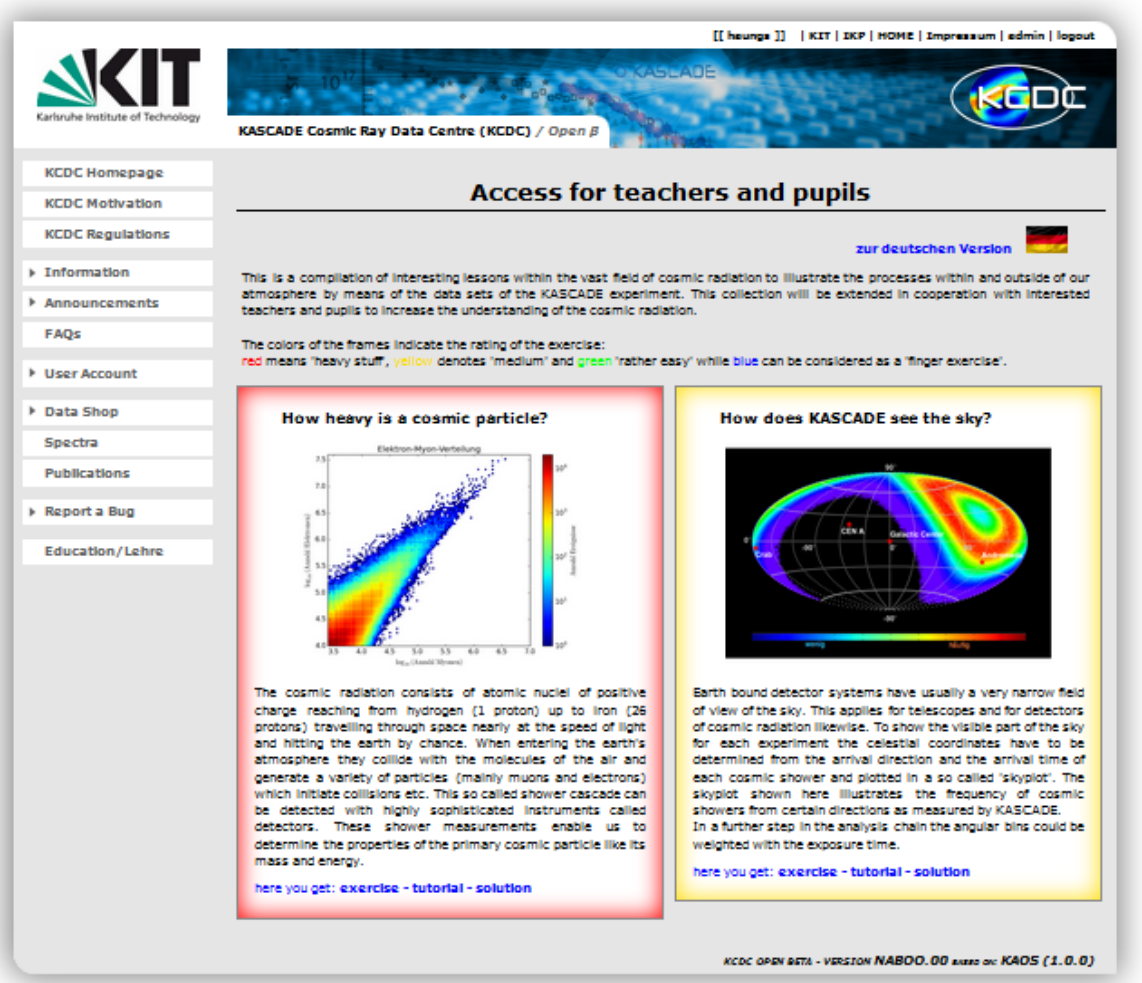
- no ready available open data licence
- free access to data and web portal
- good scientific practice for work with data
- citation of collaboration, KIT, and web portal mandatory
- free redistribution of data “as is”

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

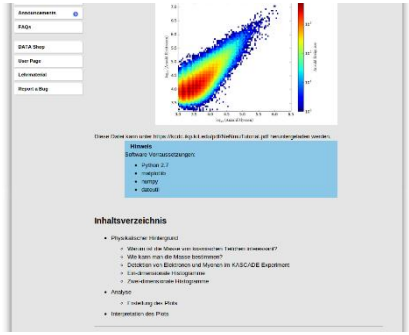


Tutorials and Teaching



- **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



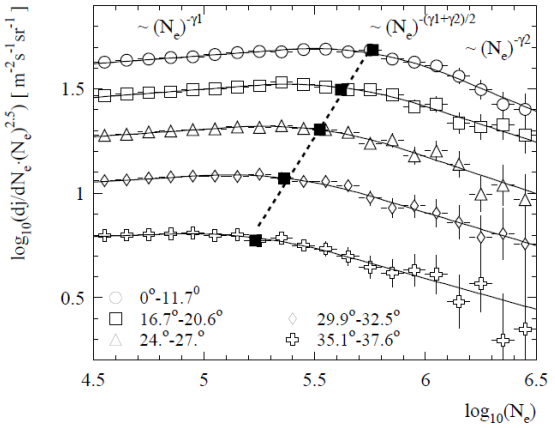
- introduction
- physics background
- step-by-step analysis
- source code example
- discussion
- interpretation
- pdf download of all

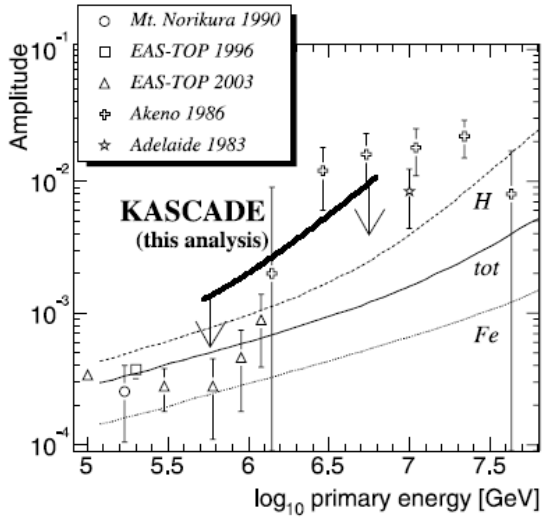
Physics with KCDC

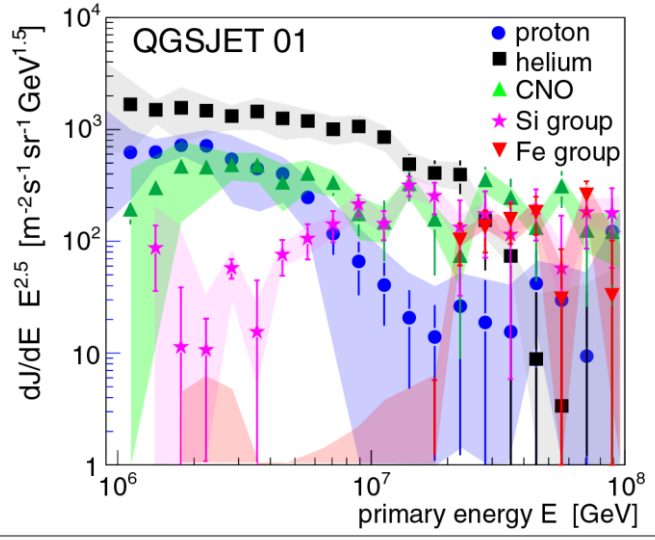
$4{,}3 \cdot 10^8$ KASCADE (and Grande) EAS events are available

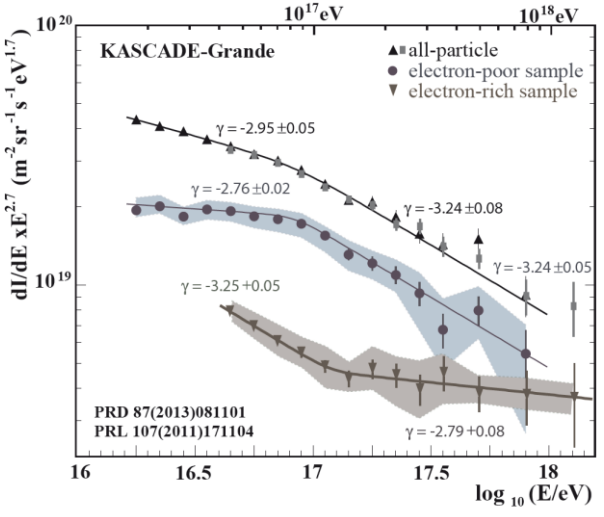
This is the data set for analysis works like

- **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.**





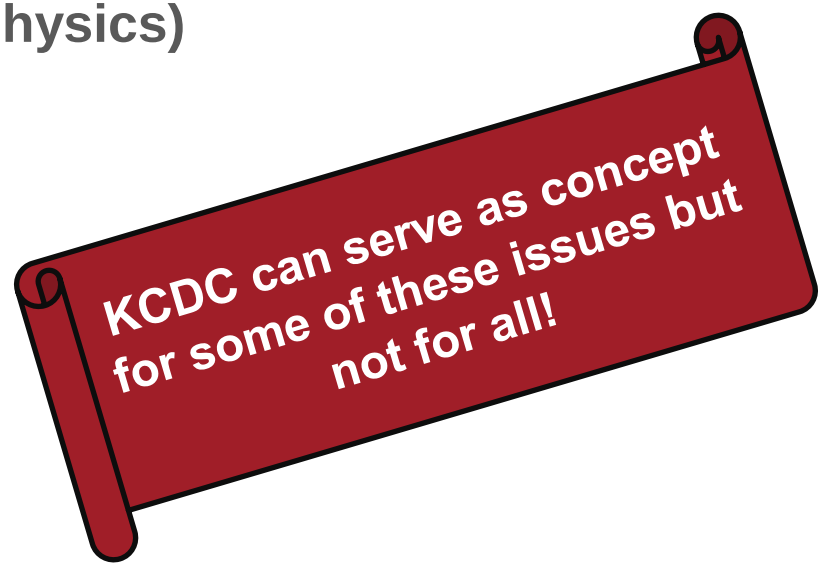




KCDC in context of Big Data Science in Astroparticle Physics

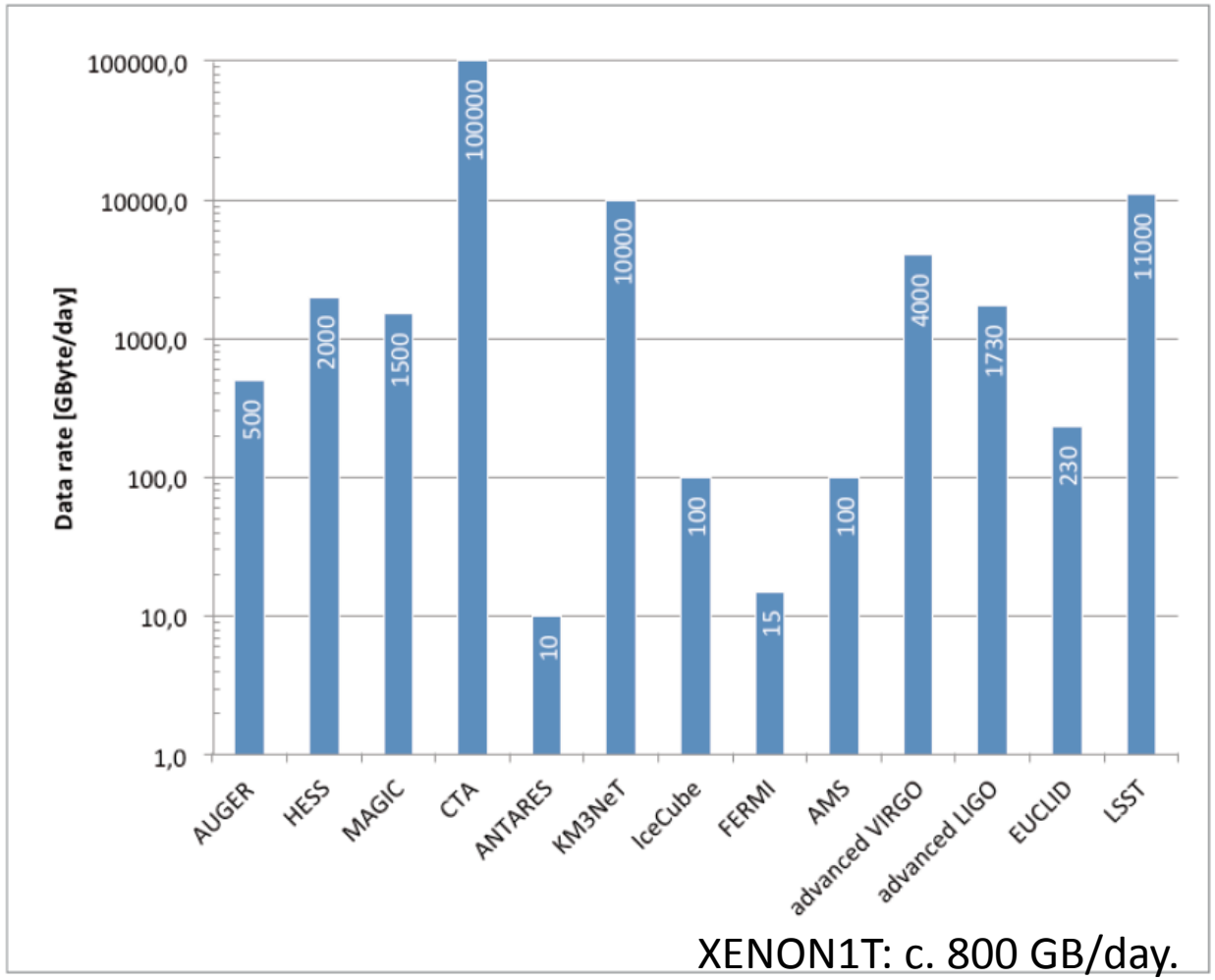
The Astroparticle Physics community needs to continue discussion on its

- 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
- …..



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

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



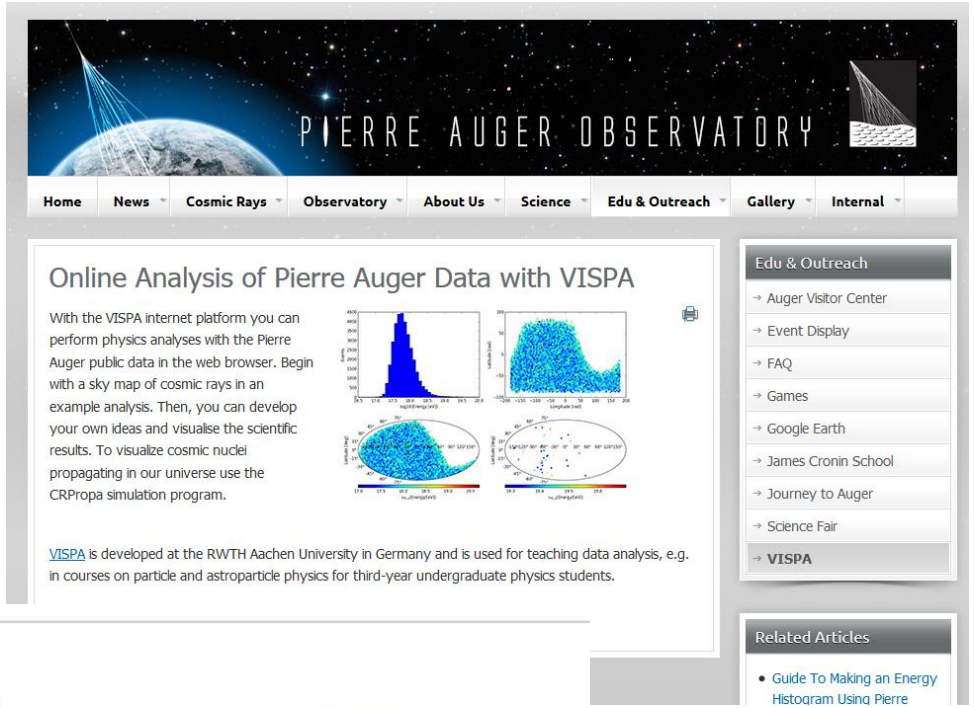
Source: APPEC brochure on Computing, 2016

➔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
- ….



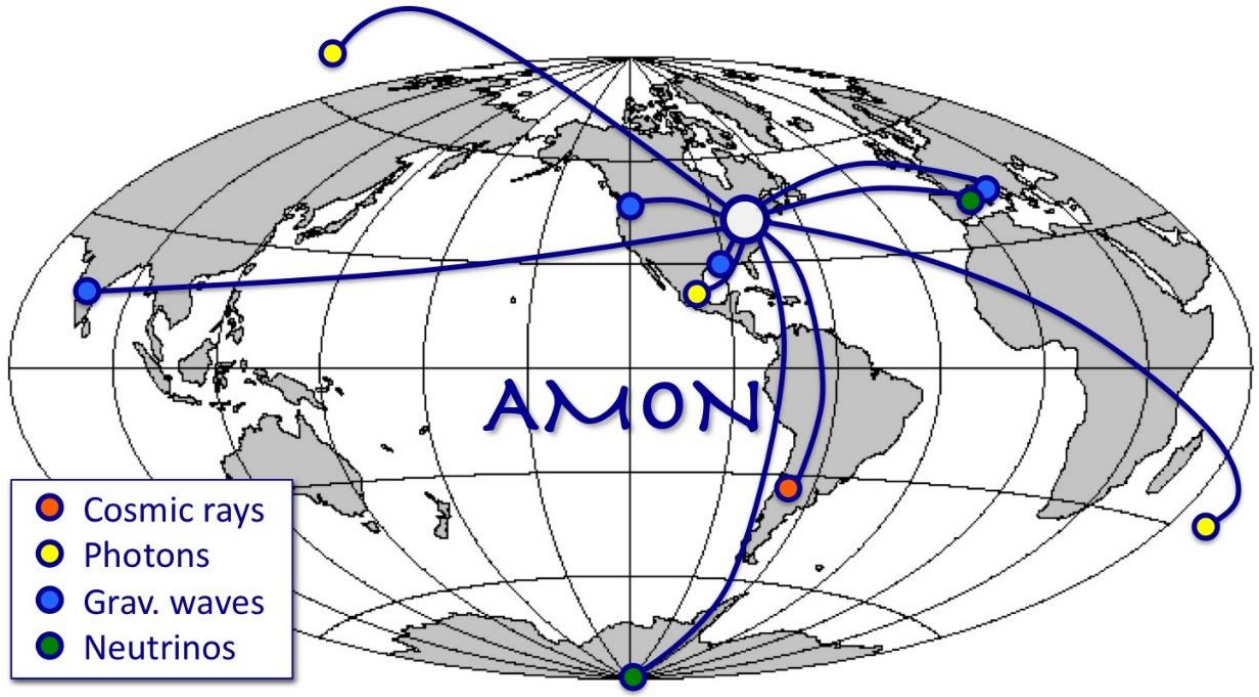
Exchange of Data / Alert systems

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

Members and Prospective Members

Observatory	Contact	Letter of Collaboration	MoU in Review	MoU Signed
ANTARES	Juergen Brunner	✓	✓	✓ MOU
Auger	Miguel Mostafa	✓	✓	✓ MOU
FACT	Adrian Biland			✓ MOU
Fermi	Julie McEnery	✓		
HAWC	Ignacio Taboada	✓	✓	✓ MOU
IceCube	Doug Cowen	✓	✓	✓ MOU
Las Cumbres Observatory Global Telescope (LCOGT)	Todd Boroson	✓	✓	✓ MOU
LIGO	Gabriela Gonzalez	✓		
Large Millimeter Telescope	Alberto Carramiñana	✓	✓	✓
MASTER	Vladimir Lipunov			✓ 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

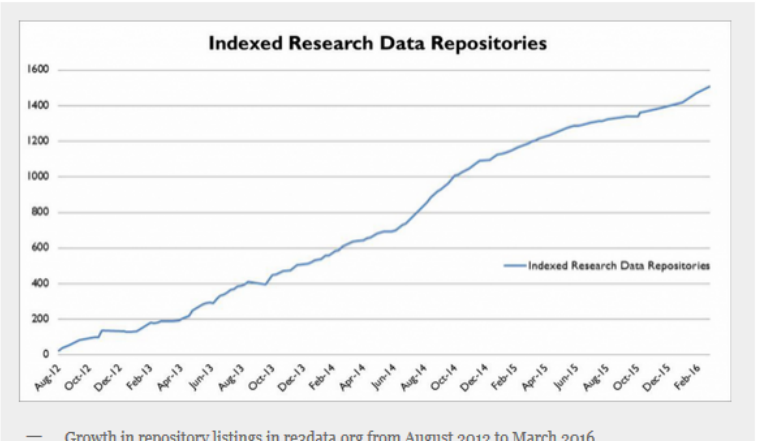
re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

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.



Growth in repository listings in re3data.org from August 2012 to March 2016

SEARCH
re3data.org search

A SERVICE BY
DataCite

PARTNERS
GFZ Helmholtz Centre Potsdam
Humboldt-Universität zu Berlin
KIT

Example Particle Physics: Data Preservation

Preservation Model		Use Case	DPHEP
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	

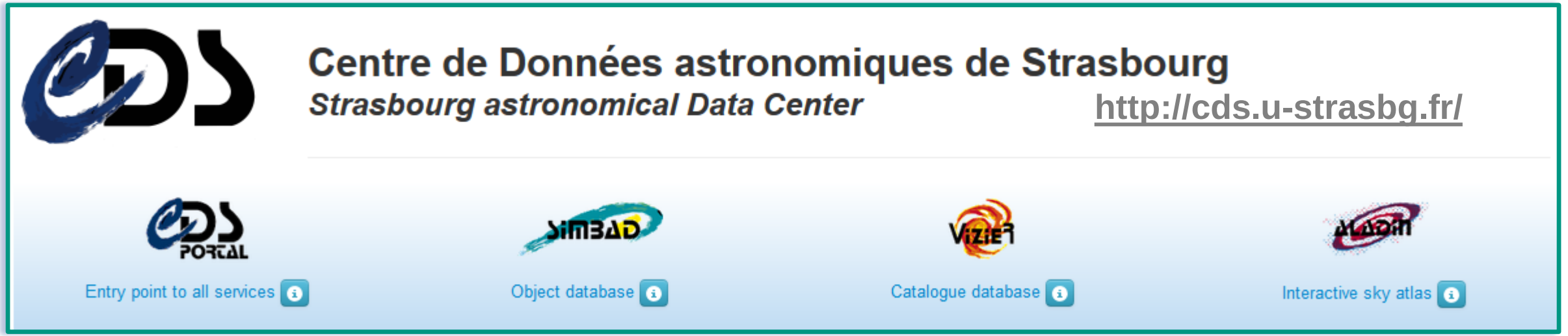
Data Preservation:

- 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



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]!

Digitale Agenda der Bundesregierung (2014-17)

https://www.digitale-agenda.de/

Handlungsfelder

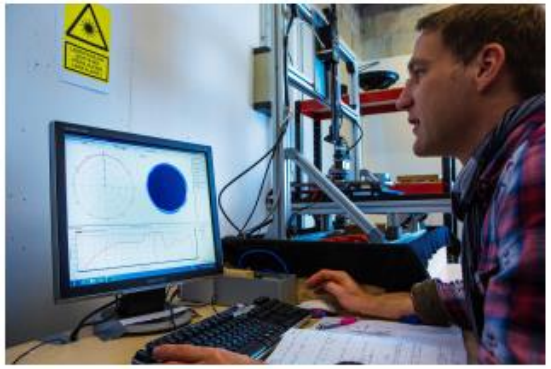
1. Digitale Infrastrukturen
2. Digitale Wirtschaft und digitales Arbeiten
3. Innovativer Staat
4. Digitale Lebenswelten in der Gesellschaft gestalten
5. Bildung, Forschung, Wissenschaft, Kultur und Medien
6. Sicherheit, Schutz und Vertrauen für Gesellschaft und Wirtschaft
7. Europäische und internationale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

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
….

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



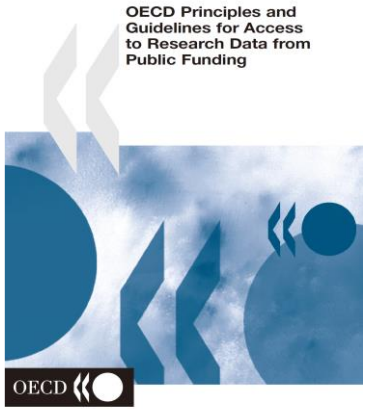
Data Center in Astroparticle Physics

- **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