Karlsruhe Institute of Technology

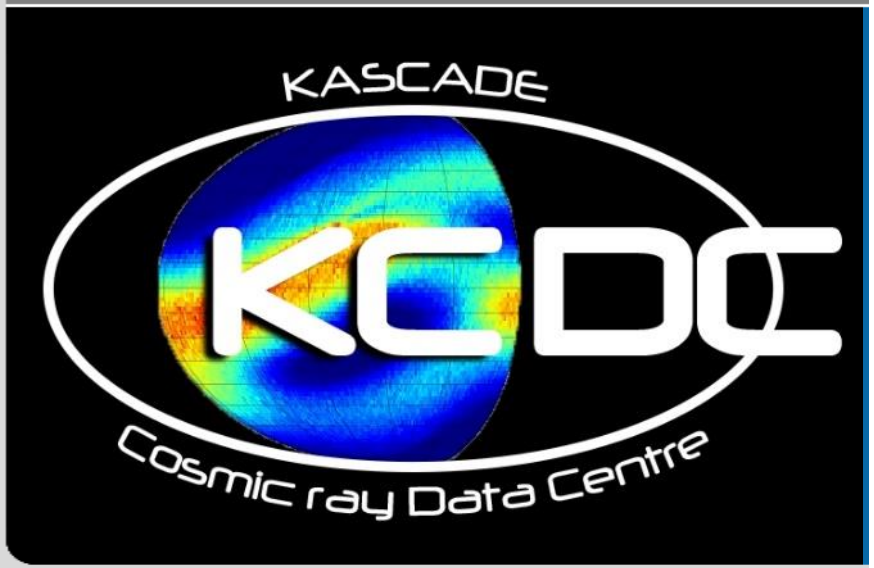
HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

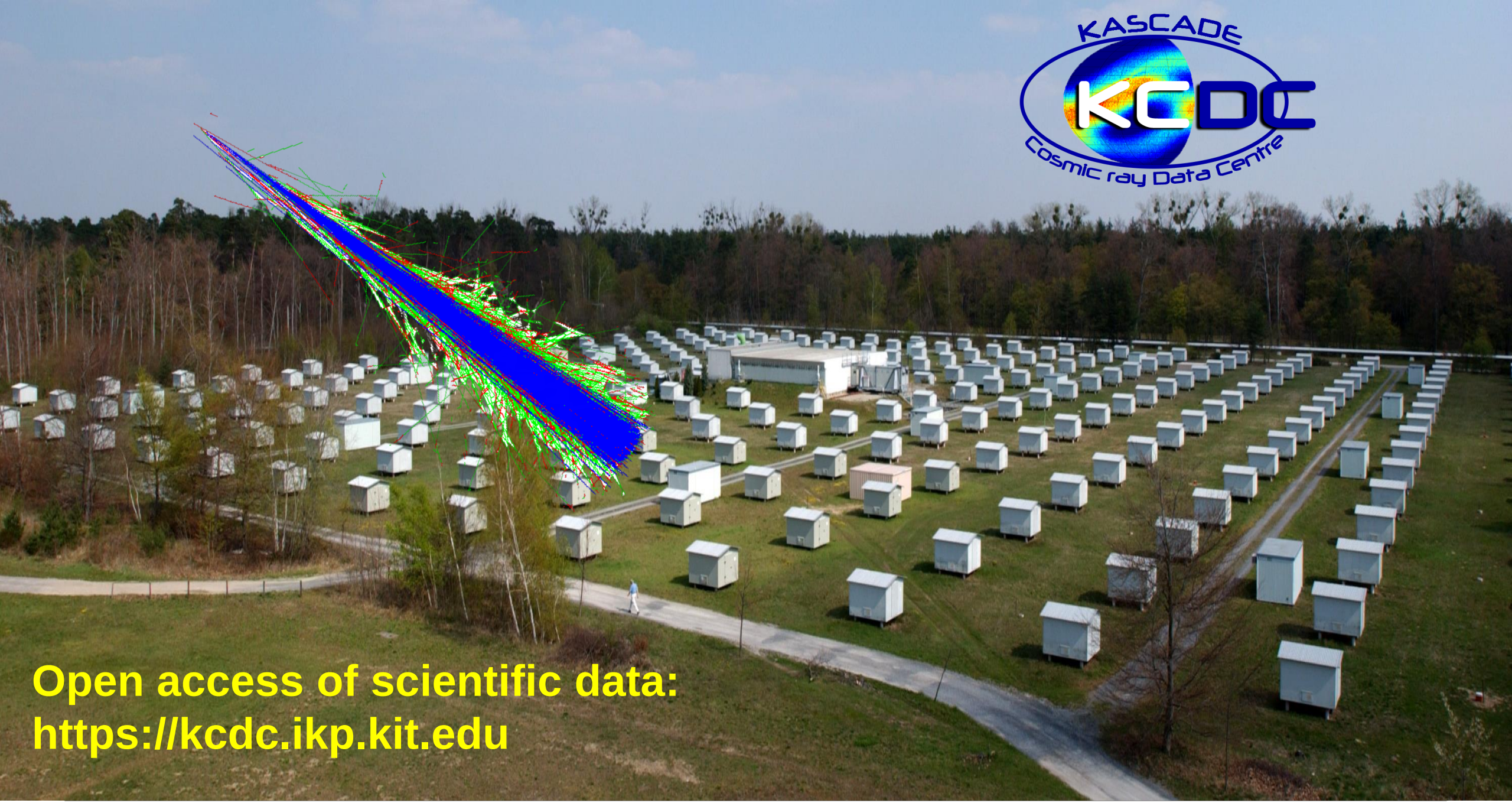
KAT. Komitee für Astro. Teilchen. Physik

KCDC and the idea of a global Analysis and Data Centre in Astroparticle Physics

KRAD/APPDS Kick-off Meeting
Moscow, 5 - 6 March 2018

Andreas Haungs

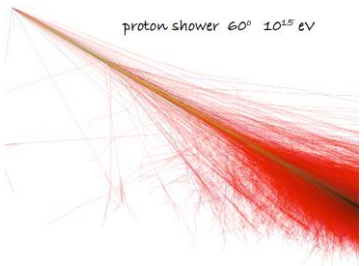



KASCADE
KCDC
Cosmic ray Data Centre
Open access of scientific data:
https://kcdc.ikp.kit.edu

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



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

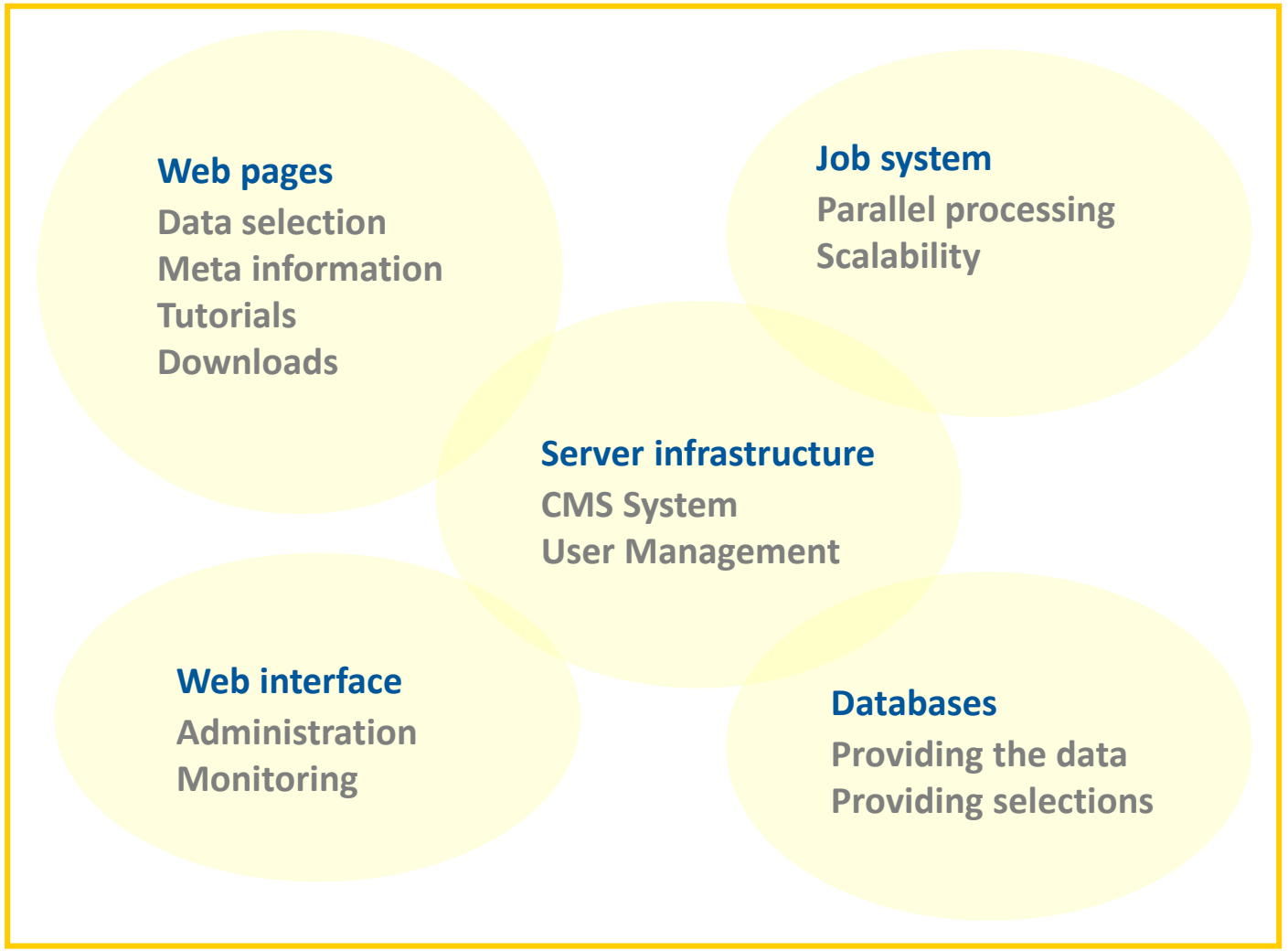
#KCDC_KIT

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

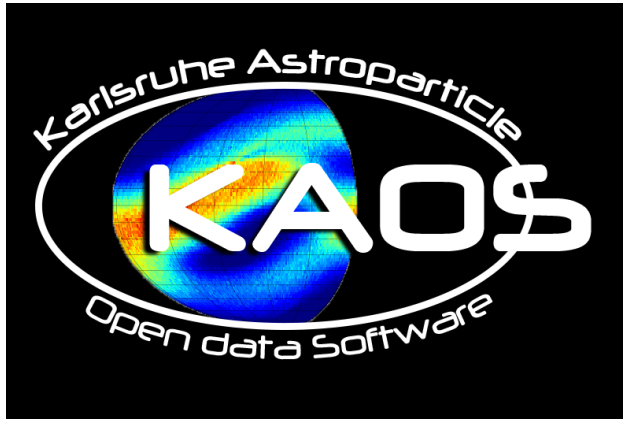
[J.Phys.Conf.Ser. 632 (2015) 012011]

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

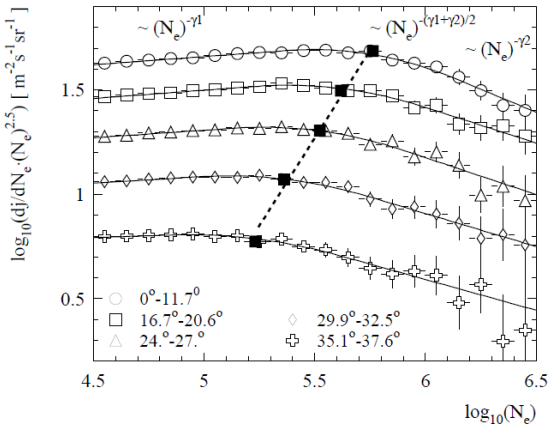
- Publication foreseen under Open Source License
- General software solution for open access to (astroparticle) data
- Following the concept of open access to research data
- Modular, flexible framework for data publication
- Good scalability (e.g. to large computing centers)
- Simple configuration via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

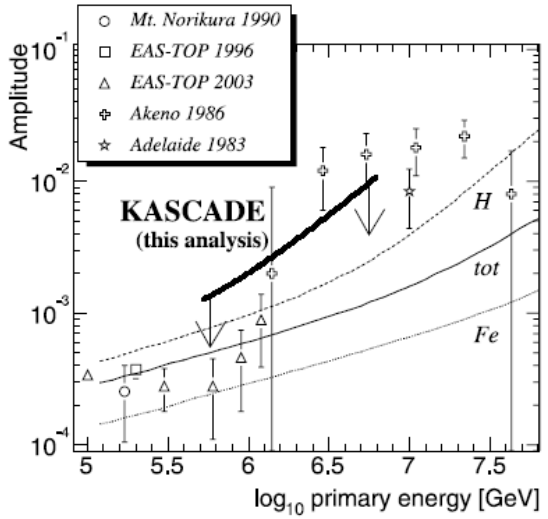
Physics with KCDC

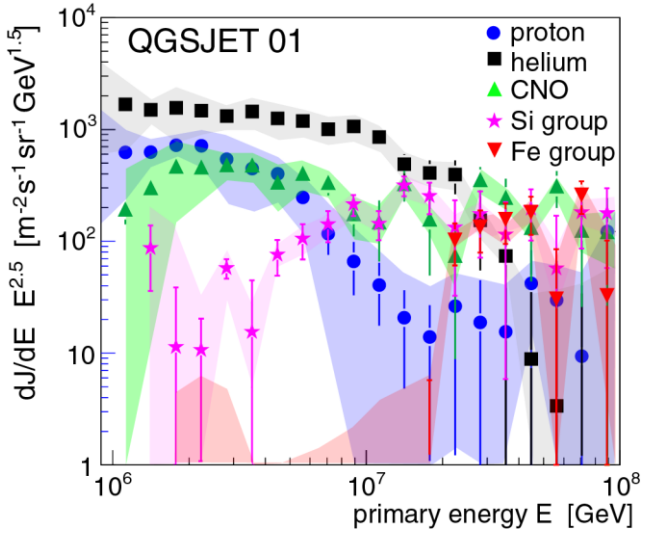
$4{,}3 \cdot 10^8$ KASCADE (and Grande) EAS events are available

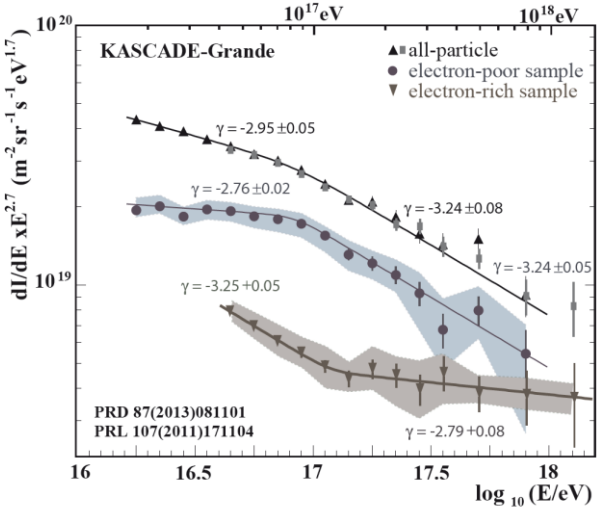
This is the data set for analysis works like

- **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.**

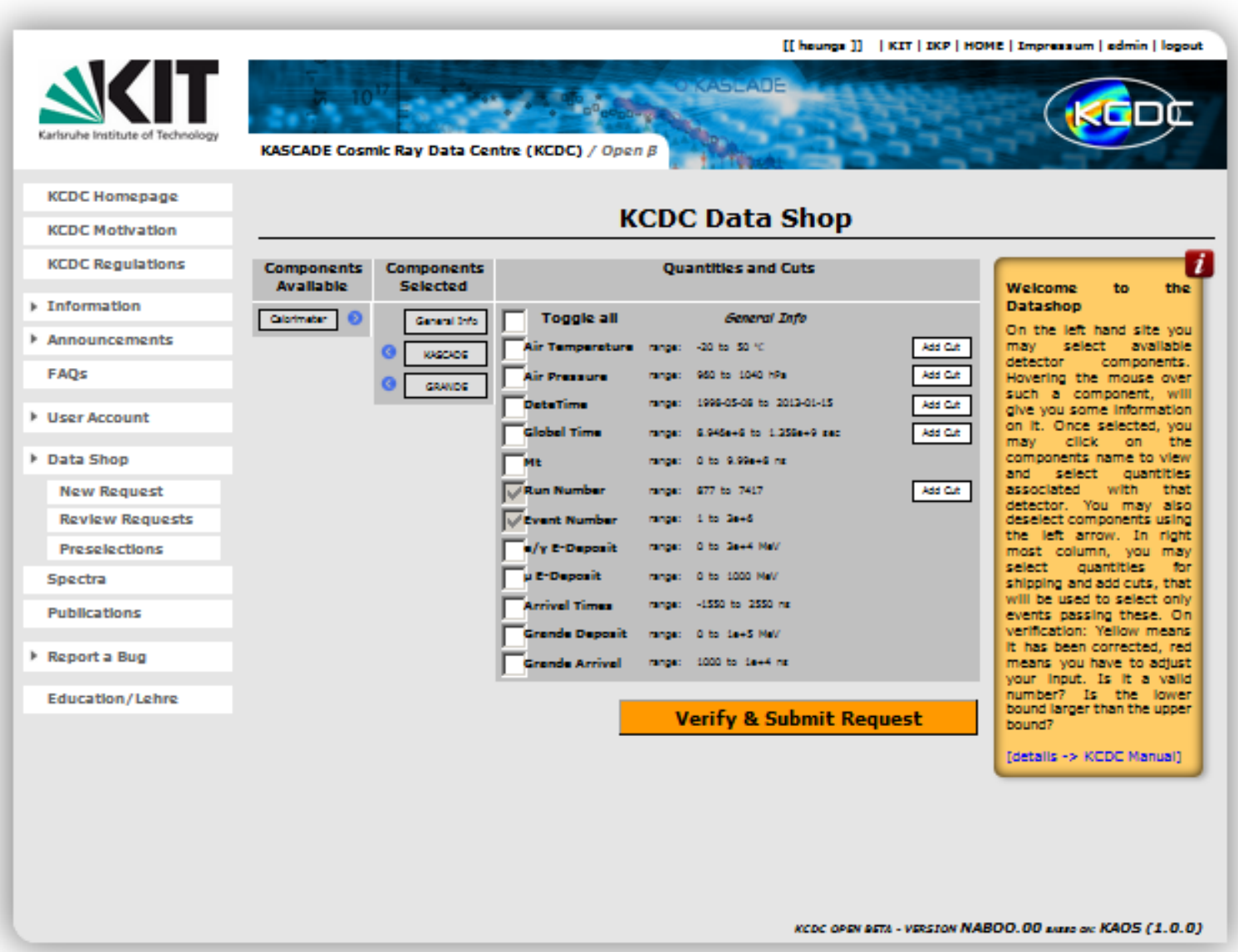








KCDC data shop



Output:

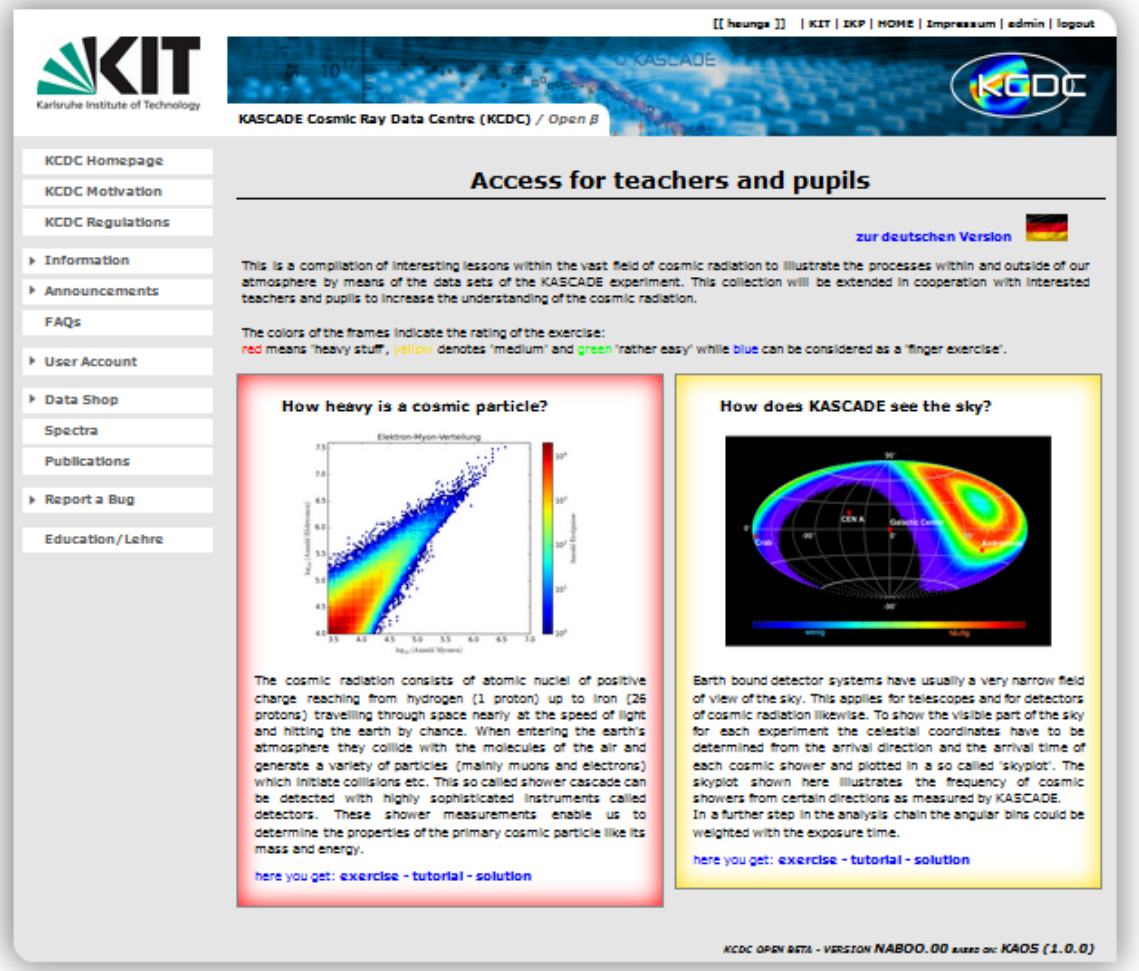
zip-archive with data, metadata, and the EULA (end user licence agreement)

Data as ASCII, ROOT and HDF5 files

Commented header give information about the content

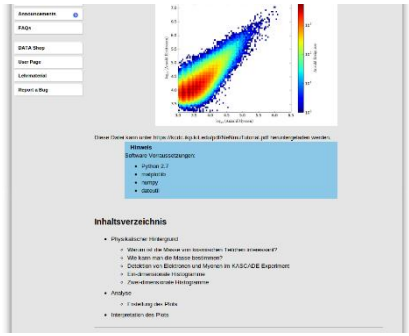
Tutorials and Teaching

#KCDC_KIT



- **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

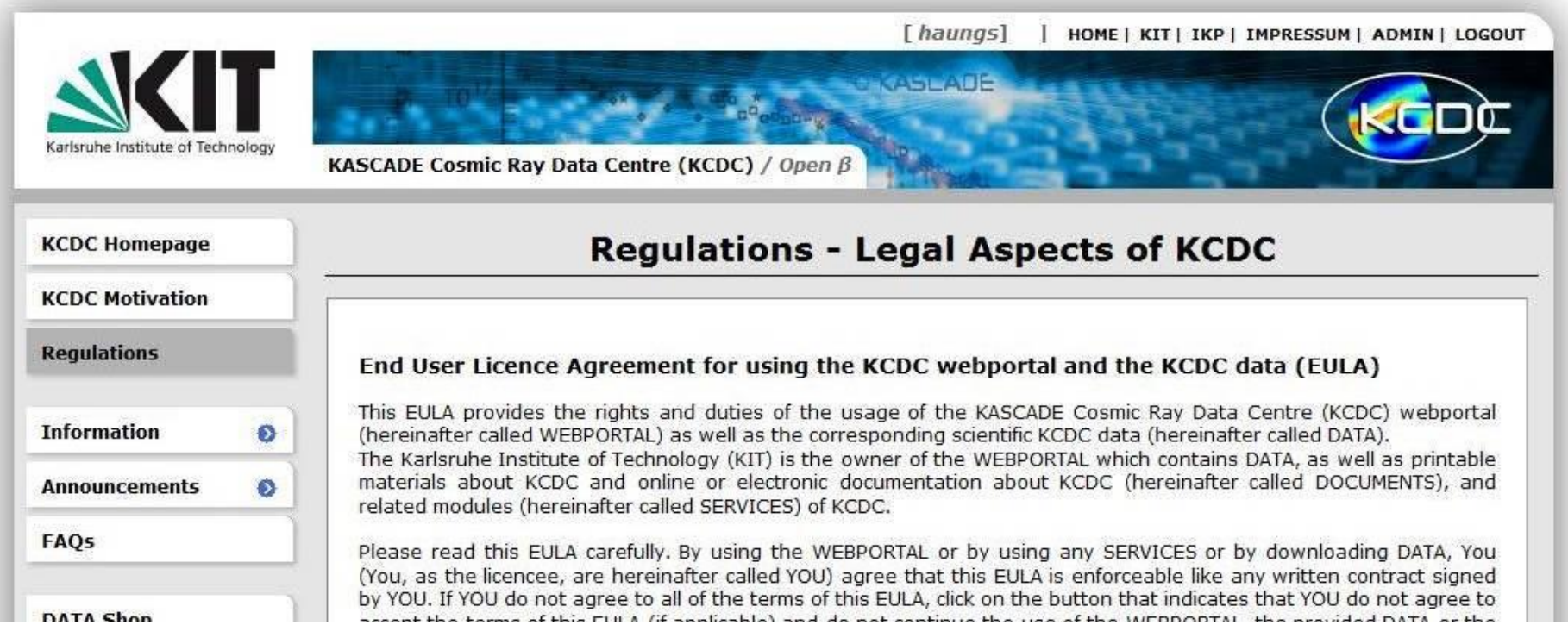
Law and Order

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”

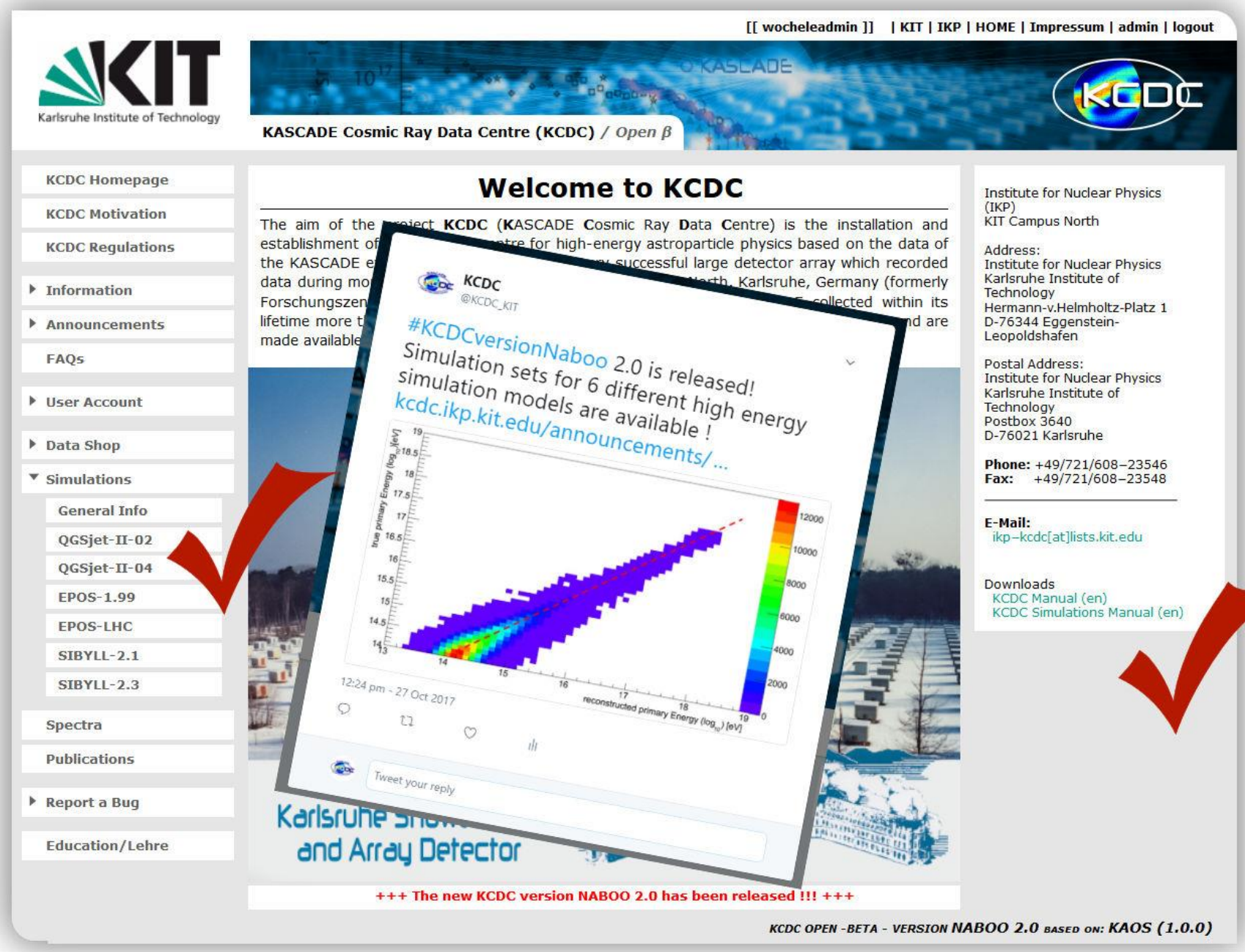
KCDC approach

- 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



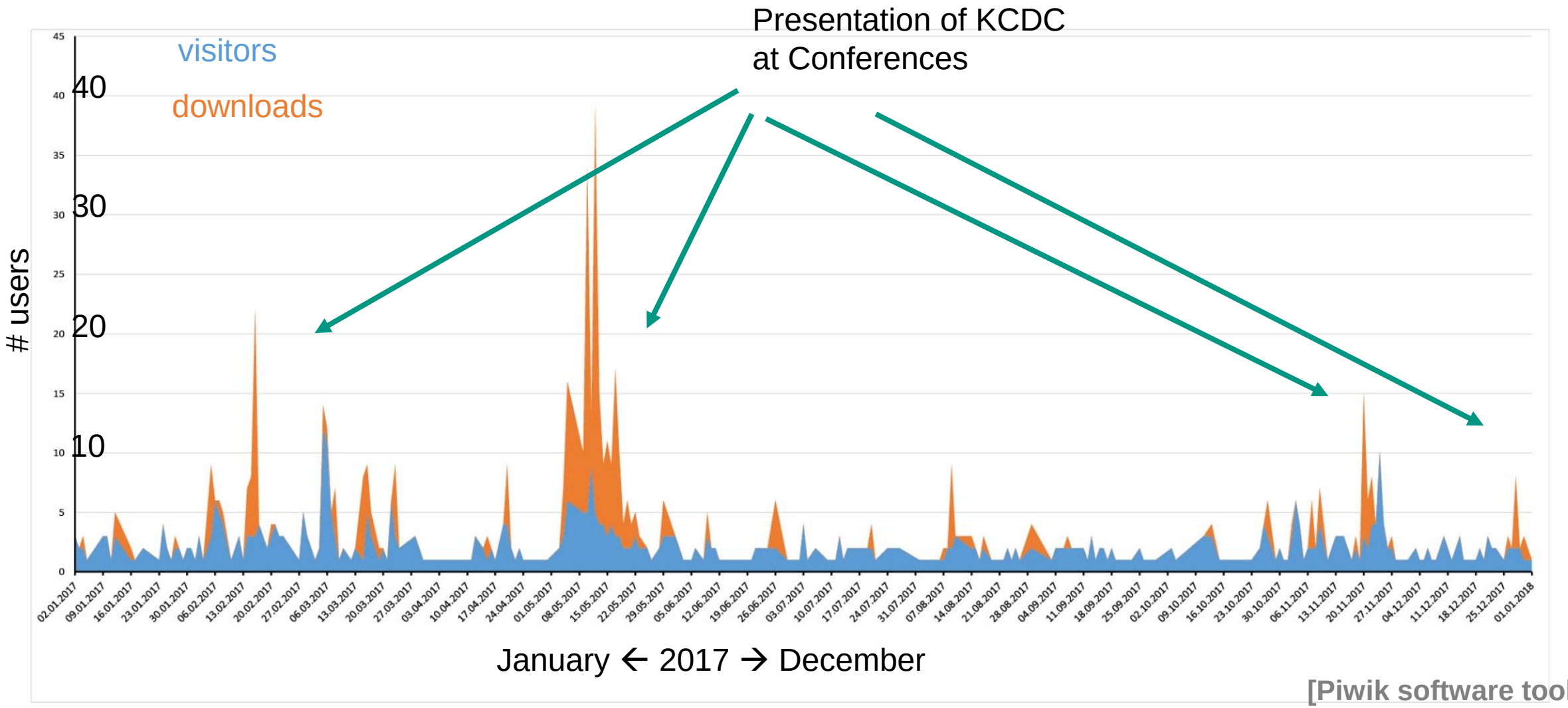
NABOO 2.0 is released!

27.10.2017



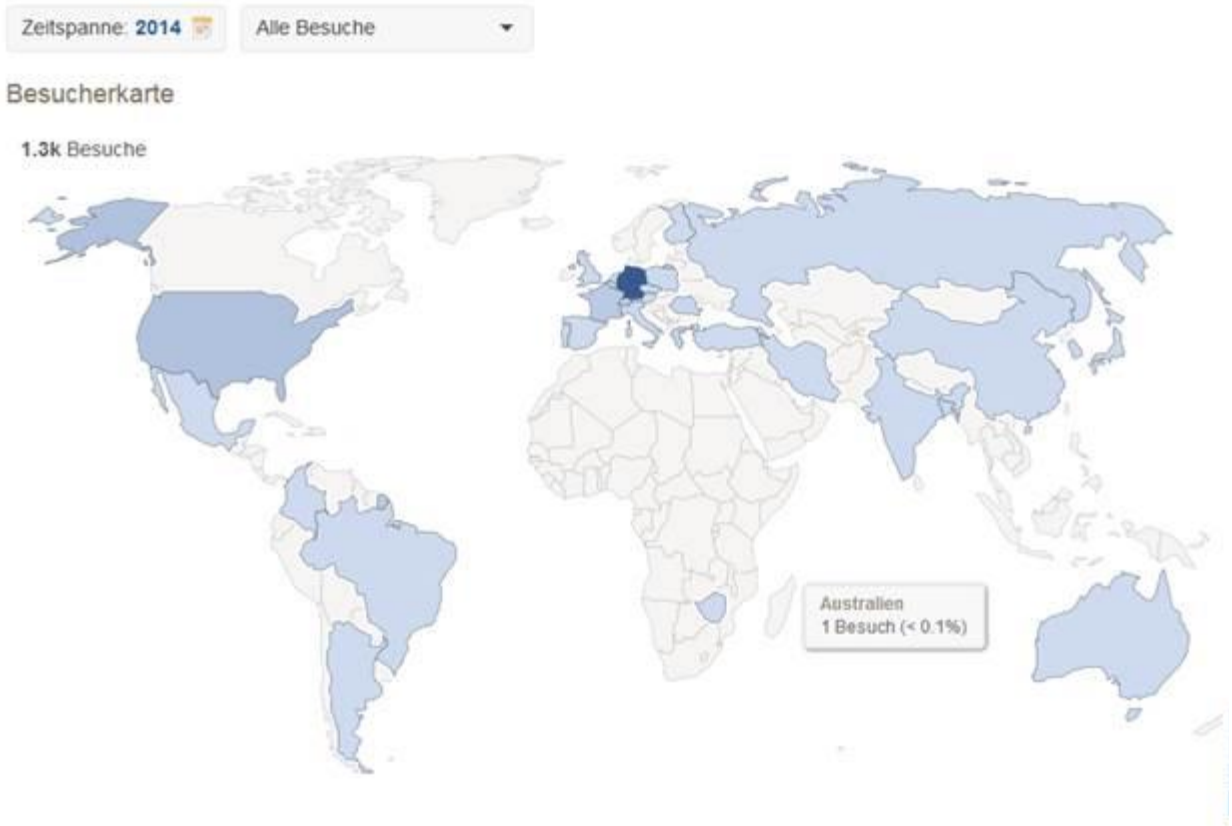
HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

KASCADE Cosmic ray Data Centre



[Piwik software tool]

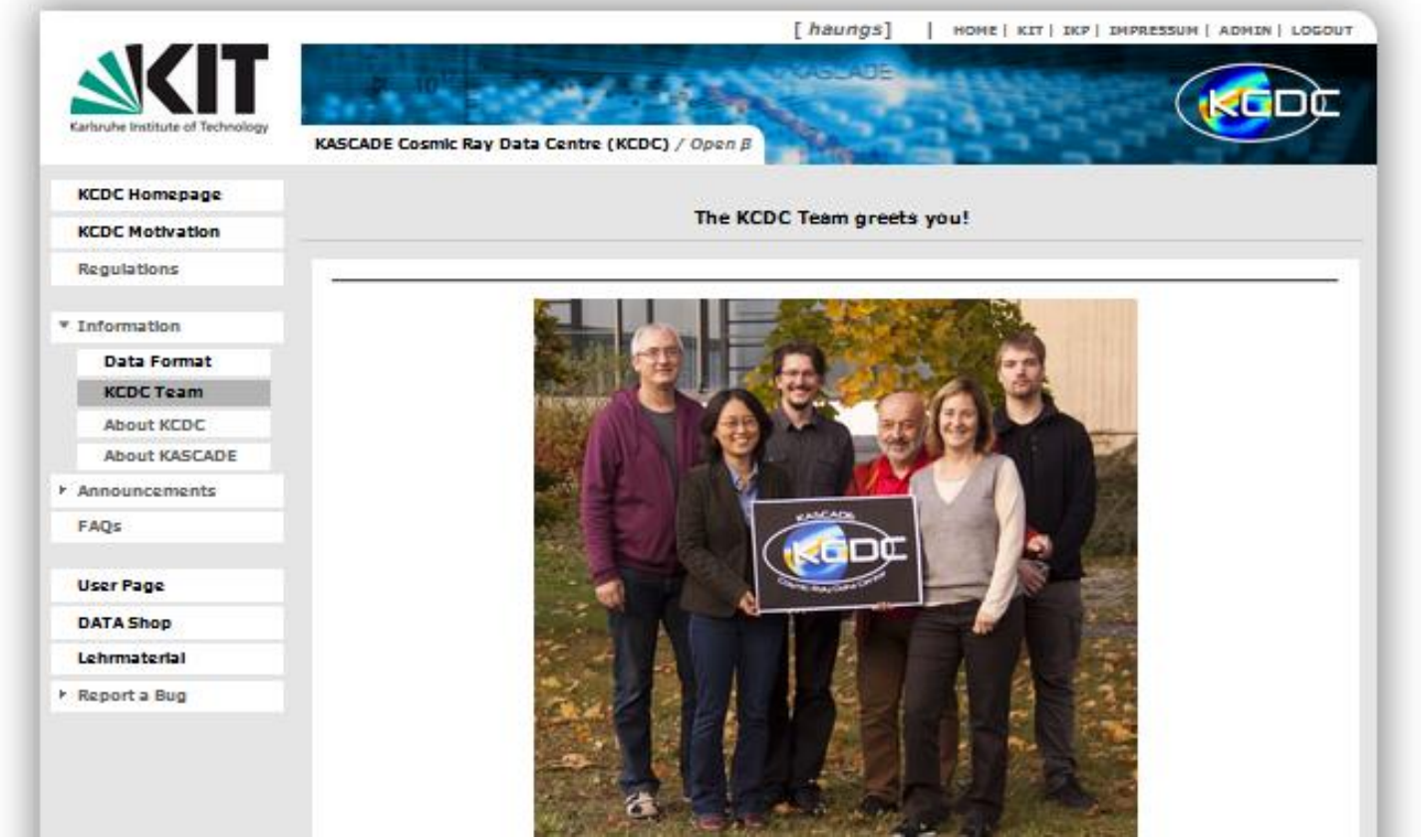
Usage



Land

Land	Besuche
Deutschland	862
USA	158
Frankreich	99
Griechenland	52
Rumänien	23
Großbritannien	12
Italien	10
unbekannt	10
Russische Föderation	9
Brasilien	6
Schweiz	6
China	5
Indien	5
Japan	4
Polen	4
Österreich	4
Argentinien	3
Finnland	3
Mexiko	3
Portugal	2
Spanien	2
Armenien	1
Australien	1
Bangladesch	1
Belgien	1

https://kcdc.ikp.kit.edu

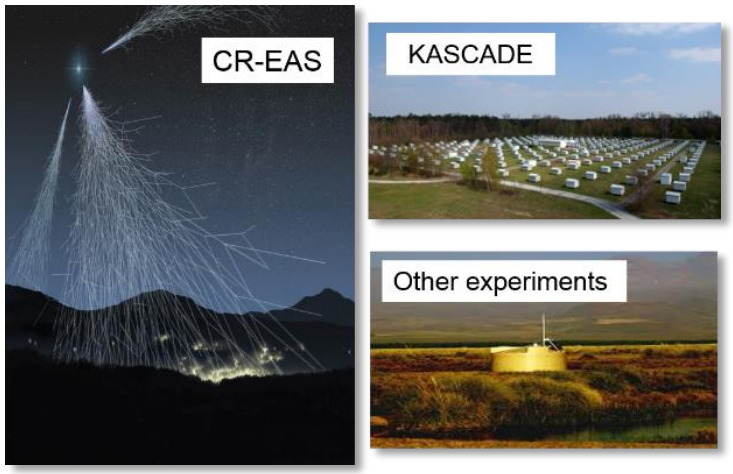


Thanks the support of the KASCADE Collaboration

Towards a (global) Analysis & Data Centre in APP

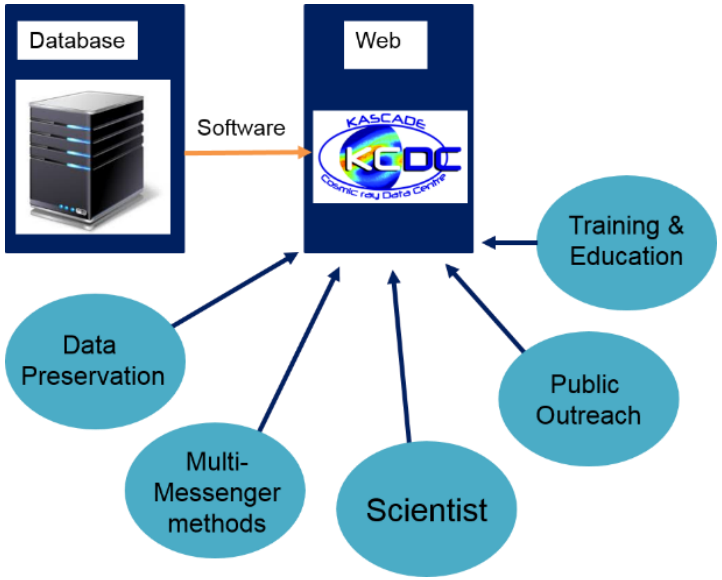
Motivation:

- 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)

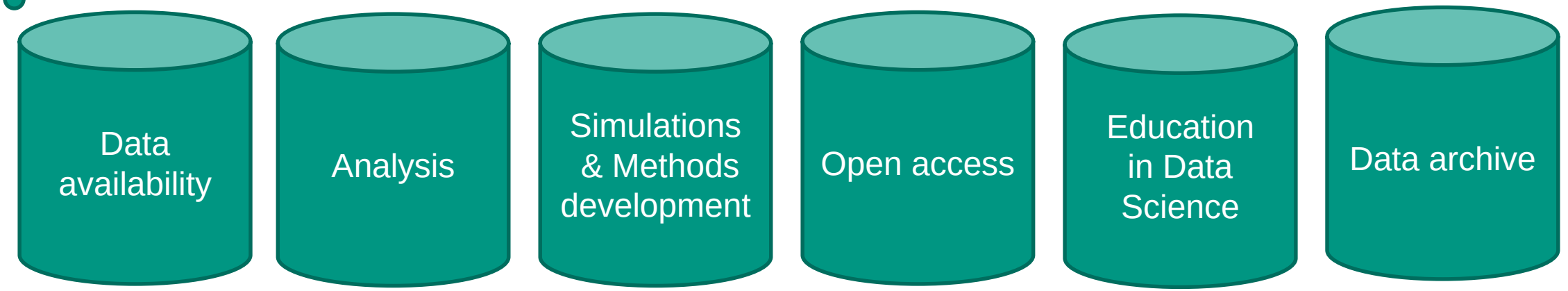


Important steps:

- 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



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

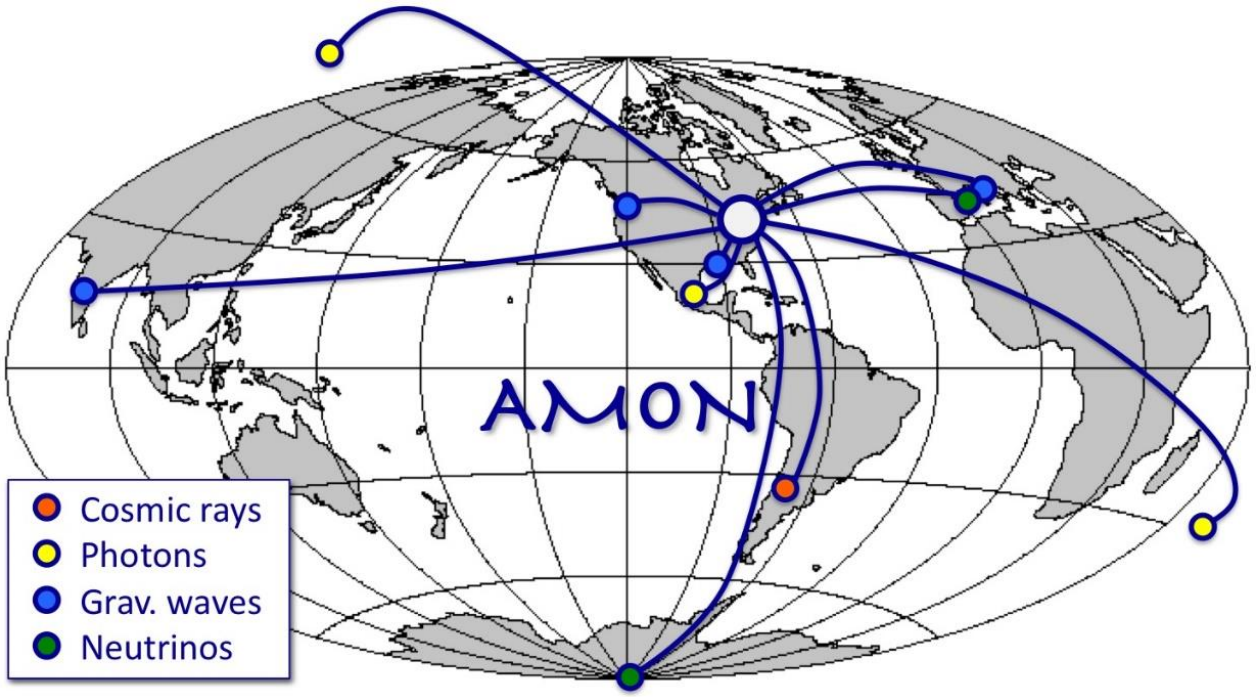
Exchange of Data / Alert systems

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

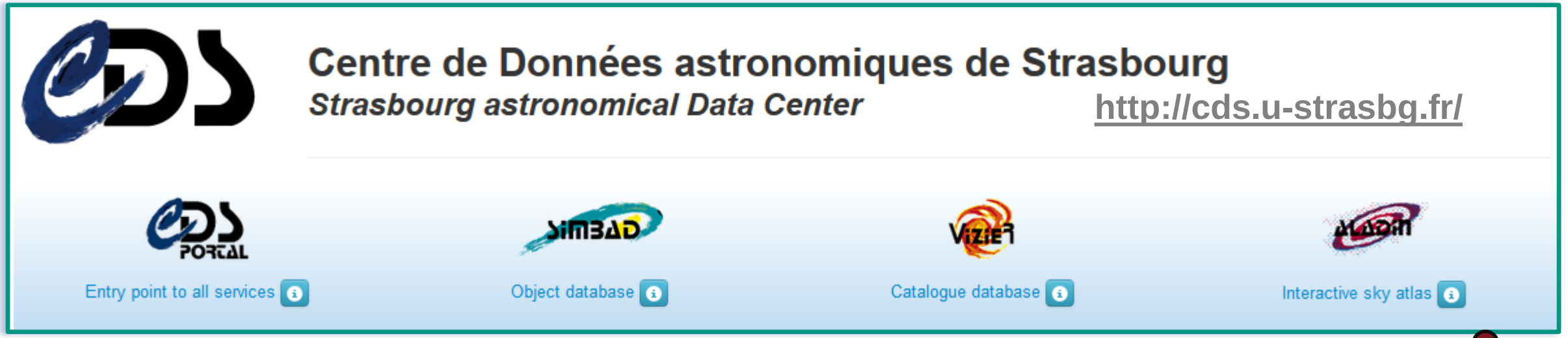
Members and Prospective Members

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.



Example Astronomy: Strasbourg astronomical Data Center



Combines many of the earlier mentioned issues:

- User Portal
- Data bases
- Tools
- Catalogues…

In Germany: GAVO in Heidelberg!

➔ What is different in astroparticle physics?

Diversity of Data, calibration, format, analysis, …

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

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

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

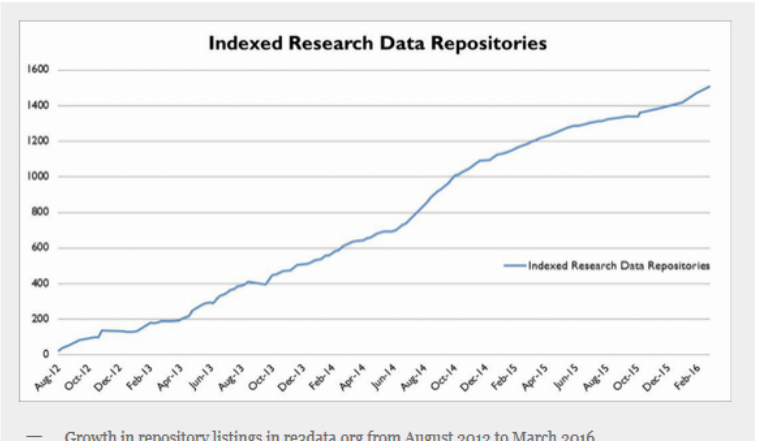
re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

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

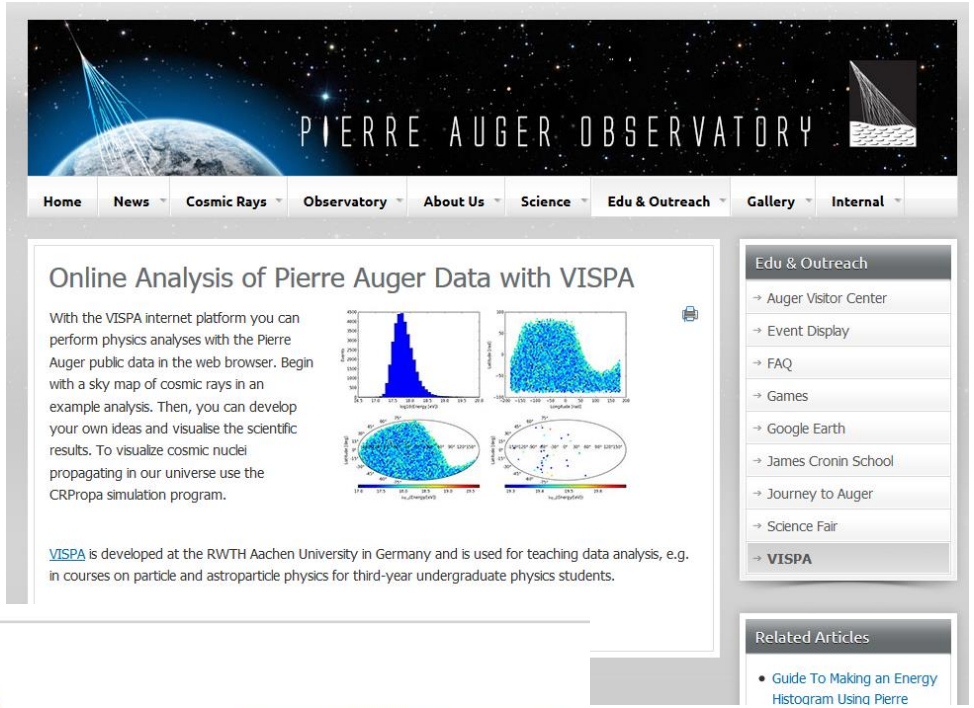
GFZ Helmholtz Centre Potsdam

Humboldt-Universität zu Berlin

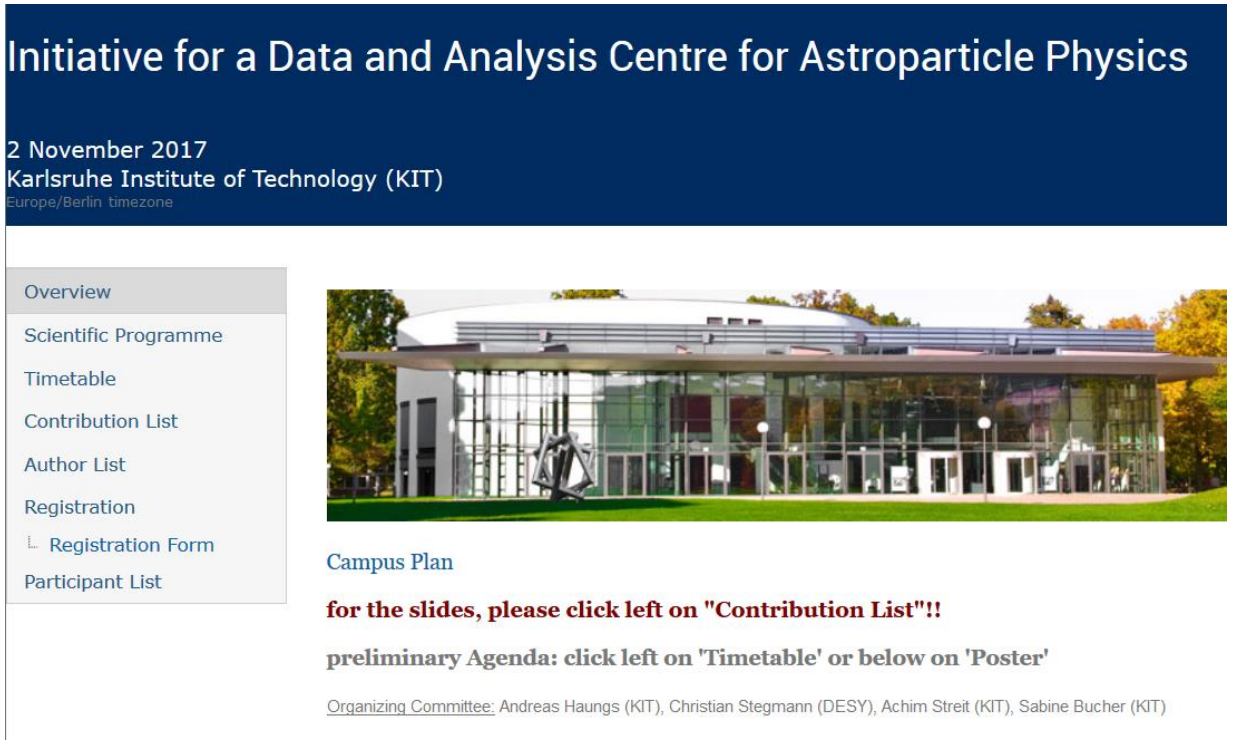
KIT

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
- (GridKa school)
- ….



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



November 2017:
40 Participants
Helmholtz + University groups

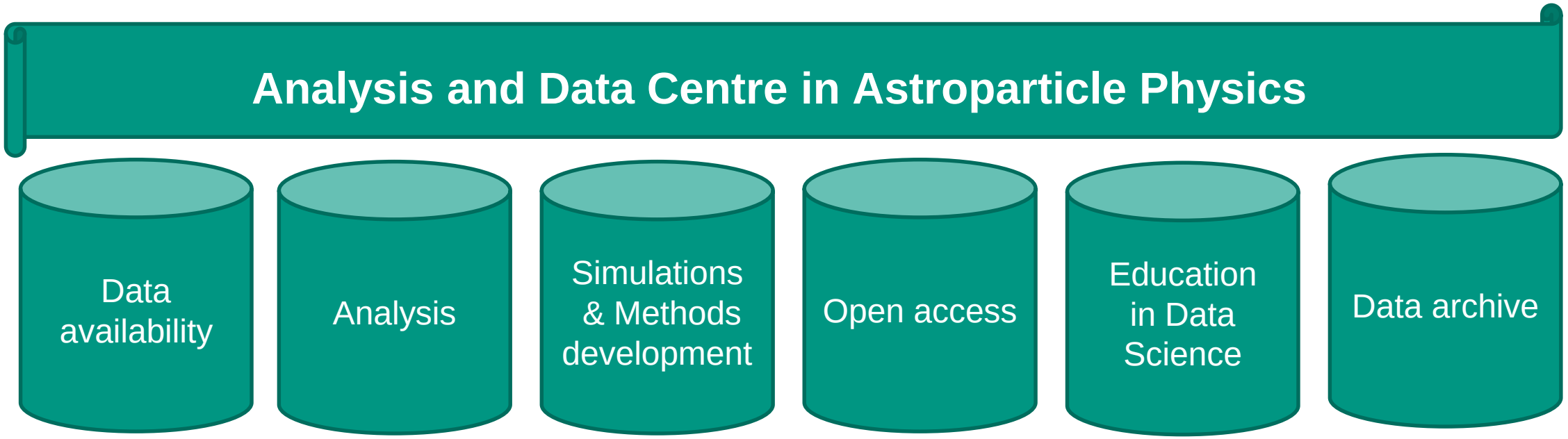
Goals:

- Data catalogues & computing resources
- Analysis & simulations
- User support & user platforms

- FAIR principles of data handling
  FINDABLE-
  ACCESSIBLE-
  INTEROPERABLE-
  REUSABLE

- Drafting white paper….

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 ☺**