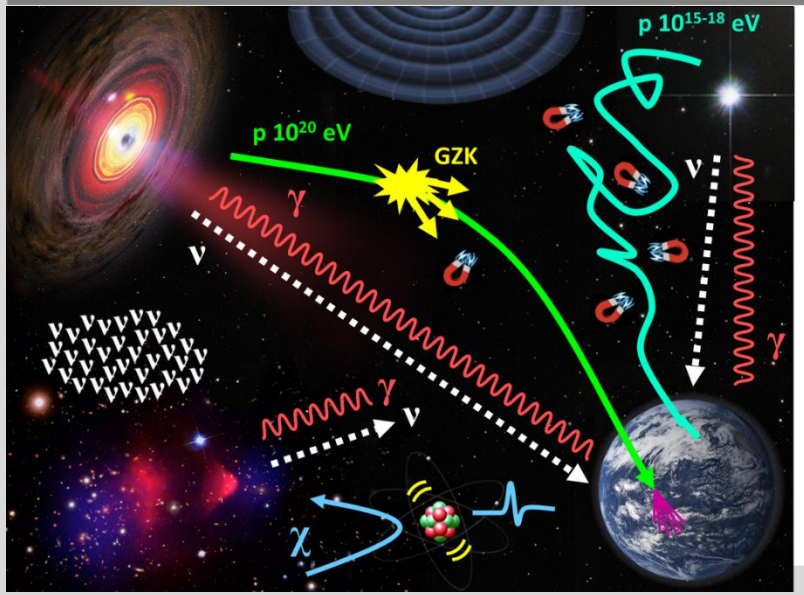
Karlsruhe Institute of Technology

HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

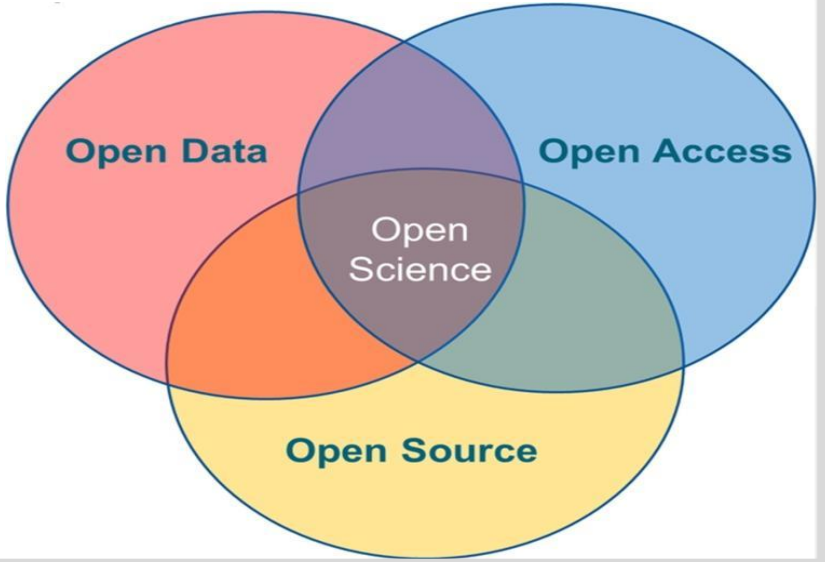
KAT. Komitee für Astro.Teilchen.Physik

# Astroparticle Physics in the Light of the Digital Agenda of the German Federal Government


Andreas Haungs & many others (German Astroparticle Physics Community)







# Recommendations of the KAT (white paper)

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

| Recommendations of the KAT |
| --- |
| The KAT emphatically emphasizes 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

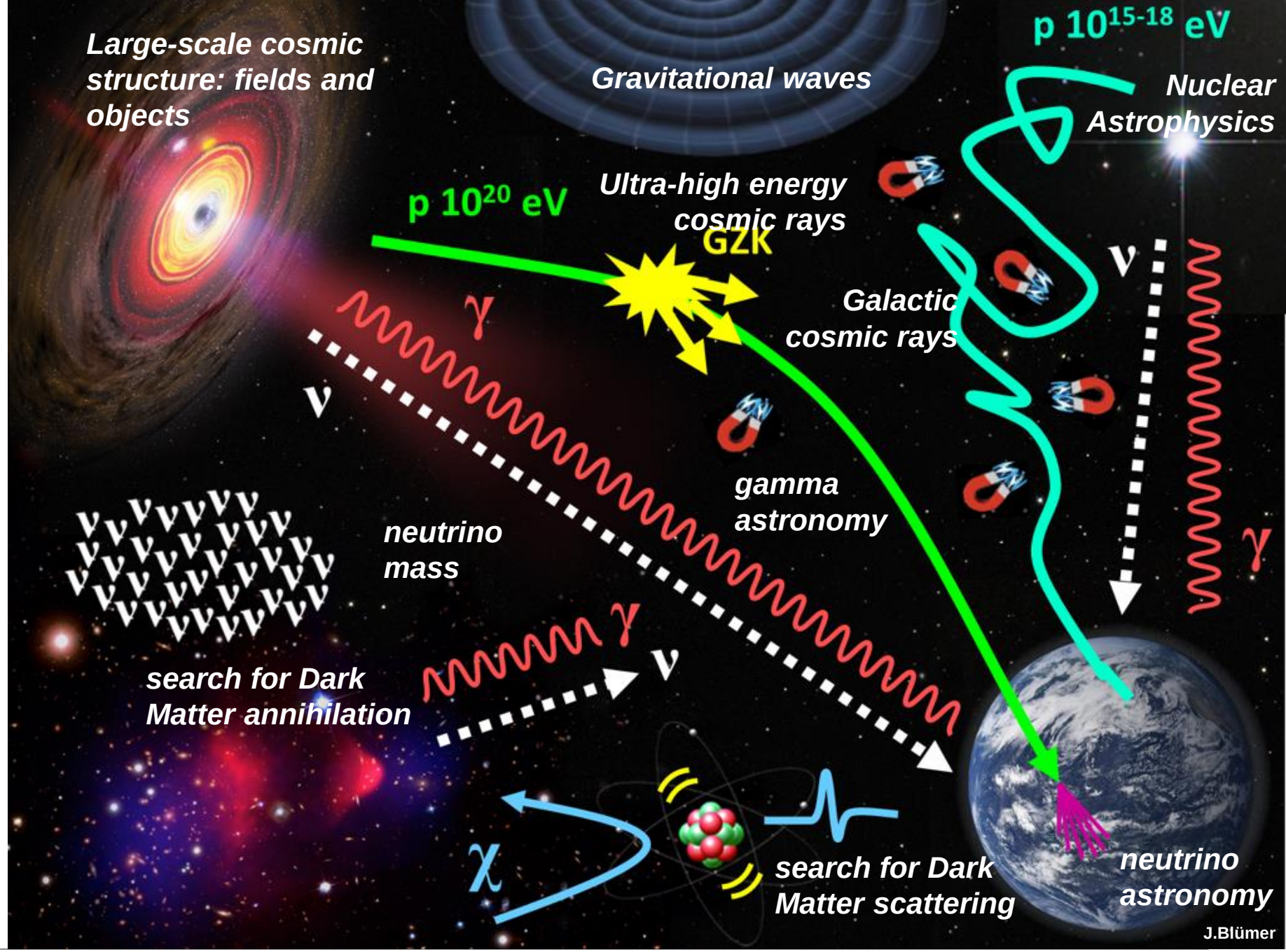


# Initiative for a (global) Analysis & Data Center in Astroparticle Physics

**Astroparticle Physics = Understanding the**

- **Multi-Messenger Universe**
- **Dark Universe**

needs an **experiment-overarching** platform!

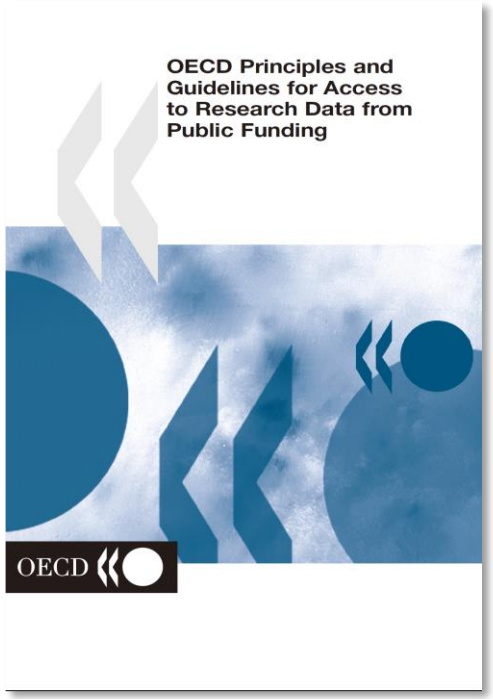


# Initiative for a (global) Analysis & Data Center in Astroparticle Physics

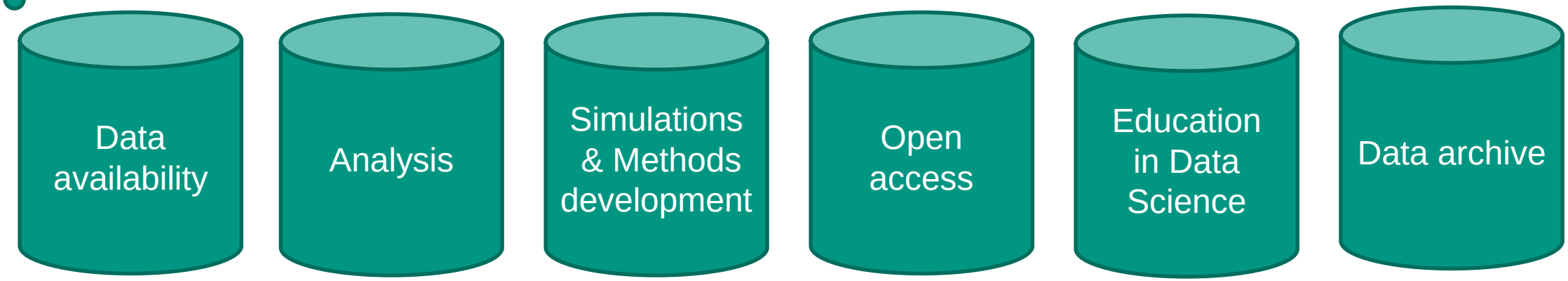
- Astroparticle Physics requests for multi-messenger analyses - this needs an **experiment-overarching** platform!

- **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!

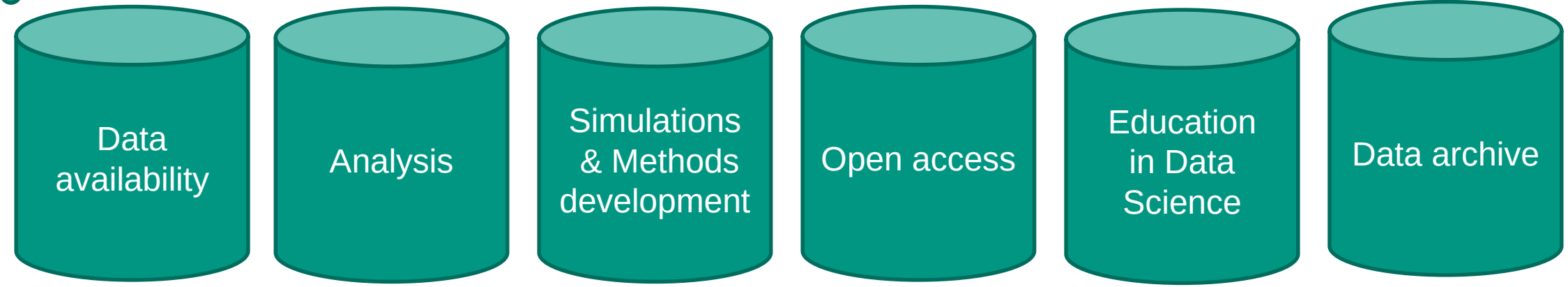
- **Education in data science:**

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



# KASCADE Cosmic ray Data Centre

- **Motivation and Idea of KCDC:**
  - 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
  - spectra
- **Pioneering work in publishing research data in astroparticle physics**

#KCDC_KIT

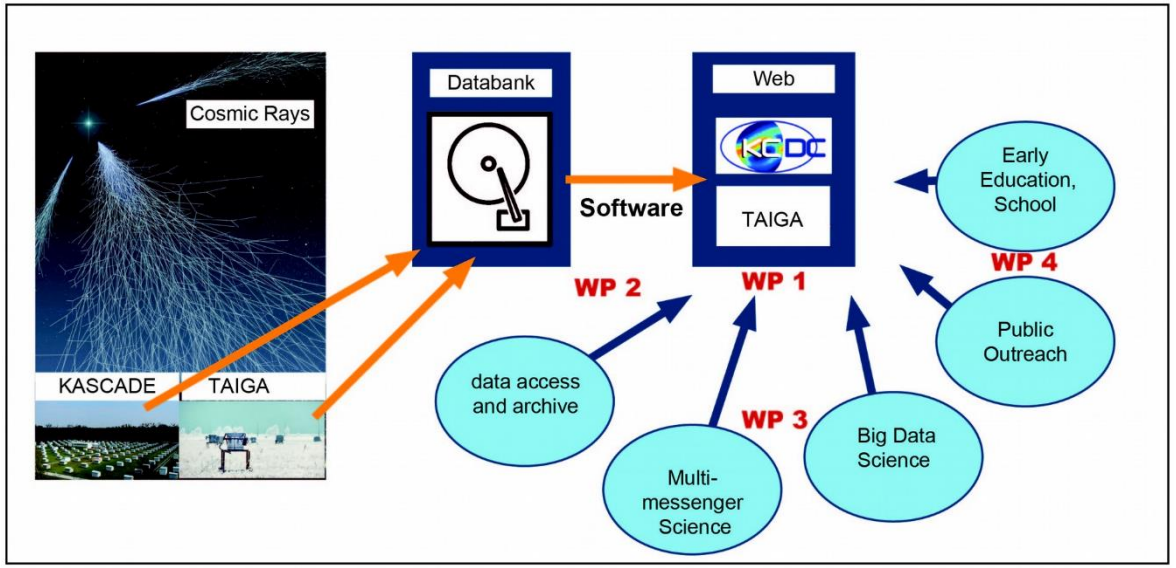

https://kcdc.ikp.kit.edu/

[J.Phys.Conf.Ser. 632 (2015) 012011]
[EPJ C (2018), in print]



# Astroparticle Data Life Cycle Initiative

- **Basics**
  - project period 2018-2020
  - funded by Helmholtz and RSF
  - 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 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

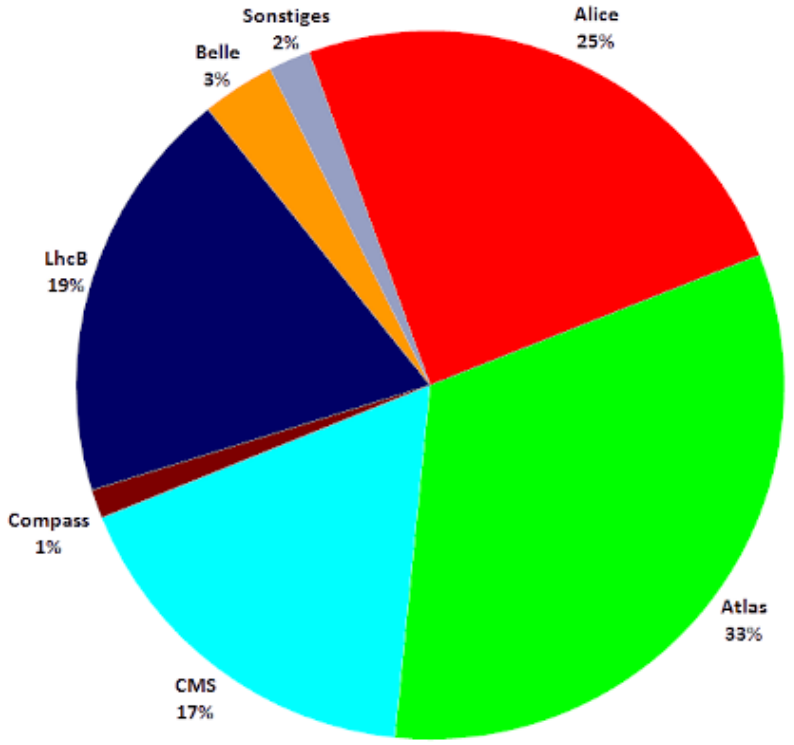
**Project is a first step in extension and generalization of KCDC**

http://astroparticle.online

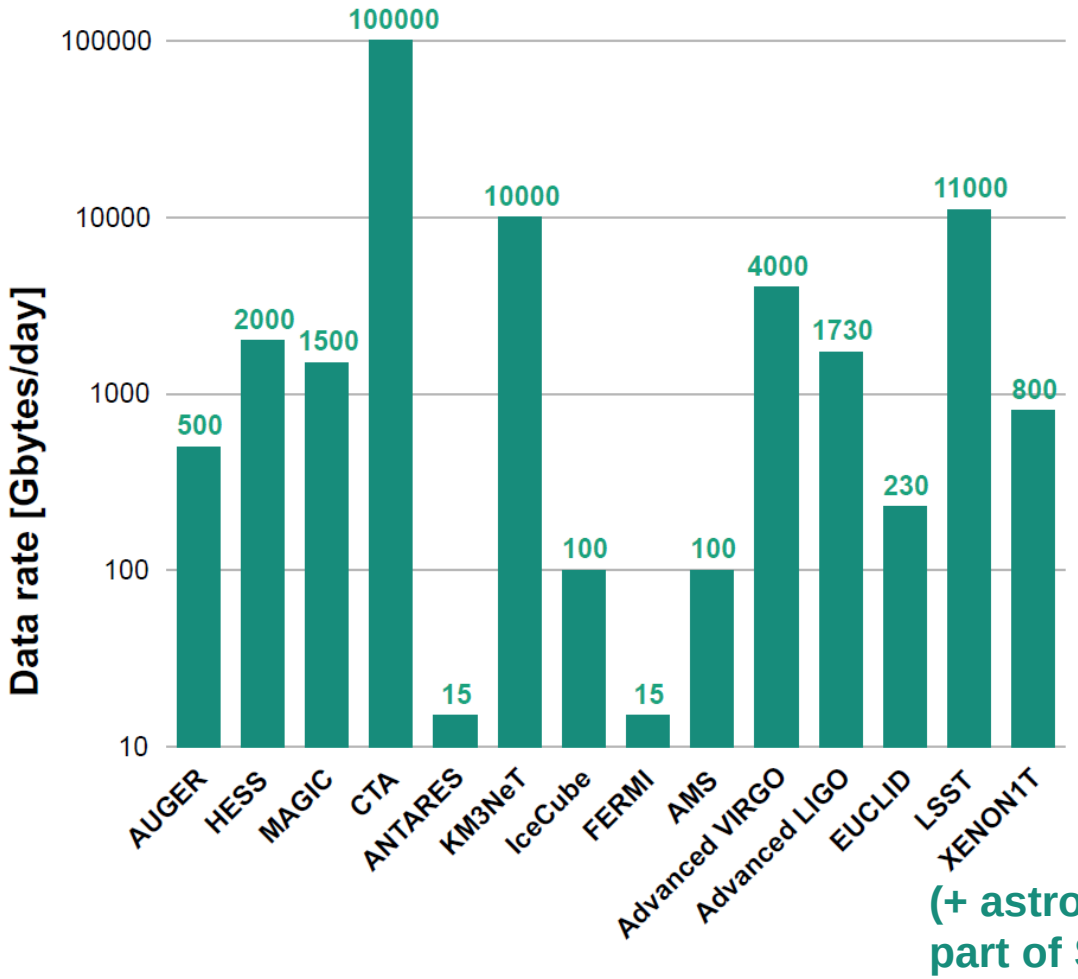
# Particle Physics: GridKa

- Central German data and computing centre for particle (and astroparticle) physics
- Tier1-centre in the world wide LHC Computing Grid
- Provides essential part of the German contribution to the LHC-Computing
- Supports non-LHC-experiments with German participation (e.g. Belle-II, Compass and Auger).

| | |
|---|---|
| Number of cores (405 kHS'06) | 22,300 |
| Number of compute jobs (2017) | 23 million |
| Number of CPU-hours delivered (2017) | 178 million |
| Disk space | 23 PB |
| Tape space (used) | 40 PB |



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




(+ astroparticle part of SKA?)

**Source: 2016 APPEC brochure on Computing: Towards a model for computing in European astroparticle physics**

## ➔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, …)?
- Is there a relation to the EOSC?

partly organized in the new facilities of astroparticle physics



# CTA Science Data Management Centre

**The Science Data Management Centre will coordinate science operations and make CTA's science products available to the worldwide community.**

- **~20 personnel will manage CTA's science coordination including software maintenance and data processing for the Observatory.**
- **CTA will generate approximately 100 petabytes (PB) of data by the year 2030.**
- **The SDMC will be located in a new building complex at DESY in Zeuthen.**
- **Provides well-established infrastructure and a powerful computing centre.**

@ DESY in Zeuthen

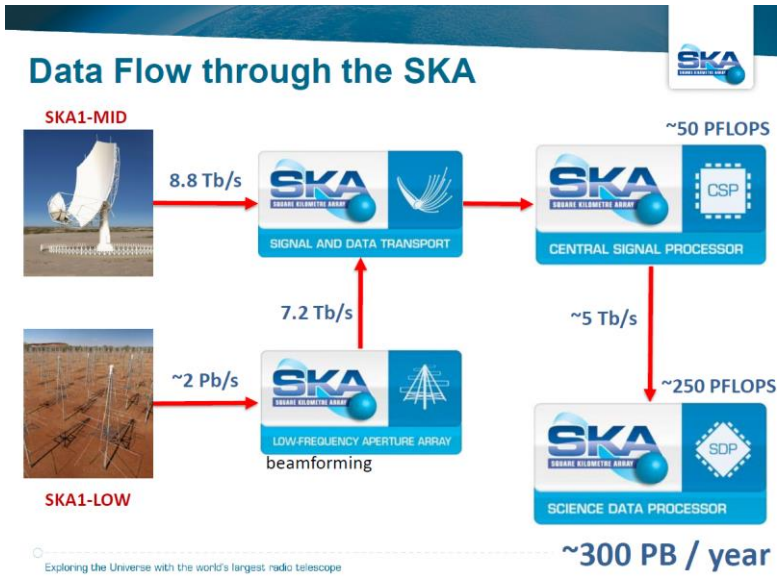
# Astronomy: Strasbourg astronomical Data Center

**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, …**

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


## Data handling of SKA




## Horizon 2020 cluster

Asterics
Astronomy ESFRI & Research Infrastructure Cluster

**ESCAPE == European Science Cluster of Astronomy and Particle Physics ESFRI's**

| | | | |
|---|---|---|---|
| CTA | ESFRI | HL-LHC | ESFRI-Landmark |
| SKA | ESFRI-Landmark | FAIR | ESFRI-Landmark |
| KM3Net | ESFRI | EGO-Virgo | |
| EST | ESFRI | JIVE | ERIC |
| ESO/E-ELT | ESFRI-Landmark | (LSST) | "observer" |

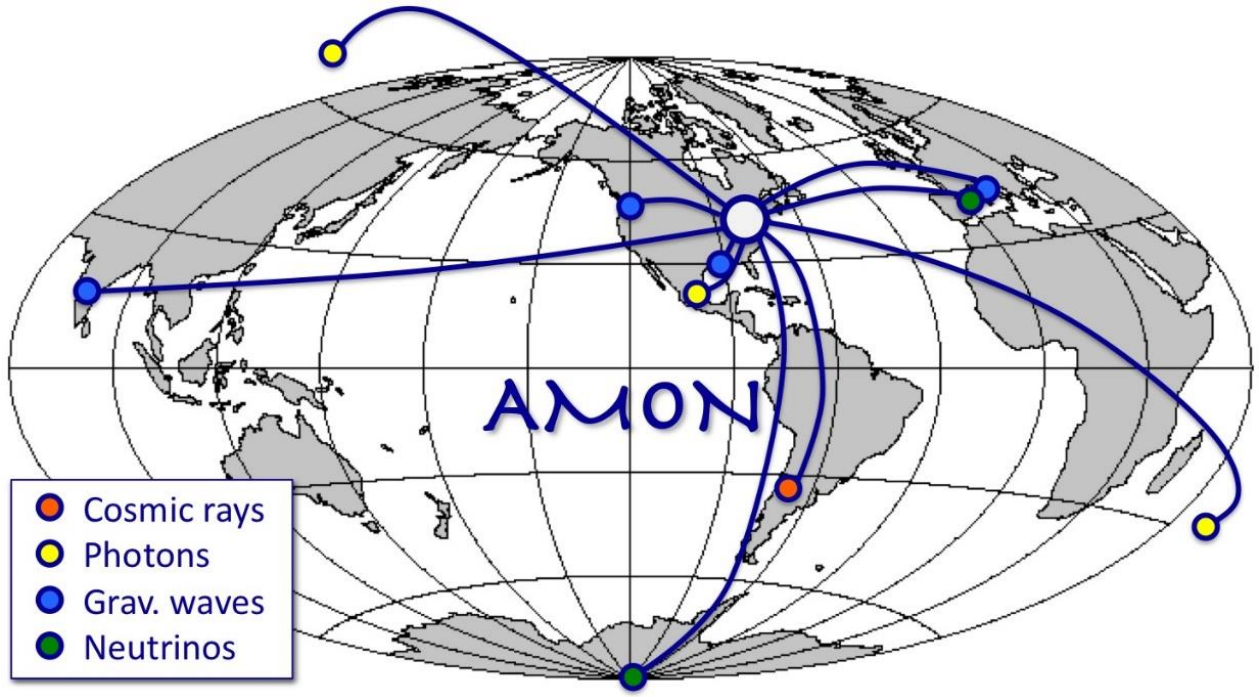
# Exchange of Data / Alert systems

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.

# Method Development & Outreach

- **VISPA to analyze (Auger) data**
  - Learning Deep Learning ☺
  - Algorithms & data analysis in own browser
  - Example analysis
  - Writing own algorithms
  - Visualizing own results
- **Dedicated lectures at Universities**
  - Deep learning in Physics Research
  - Big Data Science
- **Masterclasses**
  - Netzwerk Teilchenwelt
  - (Auger & IceCube)
- **Cosmic Days**
- **(GridKa school)**
- **....**

VISPA Internetplattform VISPA

https://vispa.physik.rwth-aachen.de/




# Data Catalogues

- **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

http://www.re3data.org/




# Support by the BMBF (I):

**Innovative Digitale Technologien für die Erforschung von Universum und Materie**

**Verbundforschung KAT-KET-KHuK (Pilotprojekt ErUM-Data)**

**Proposal** (11 Universities + 6 associated partners; coordinator T. Kuhr of LMU)

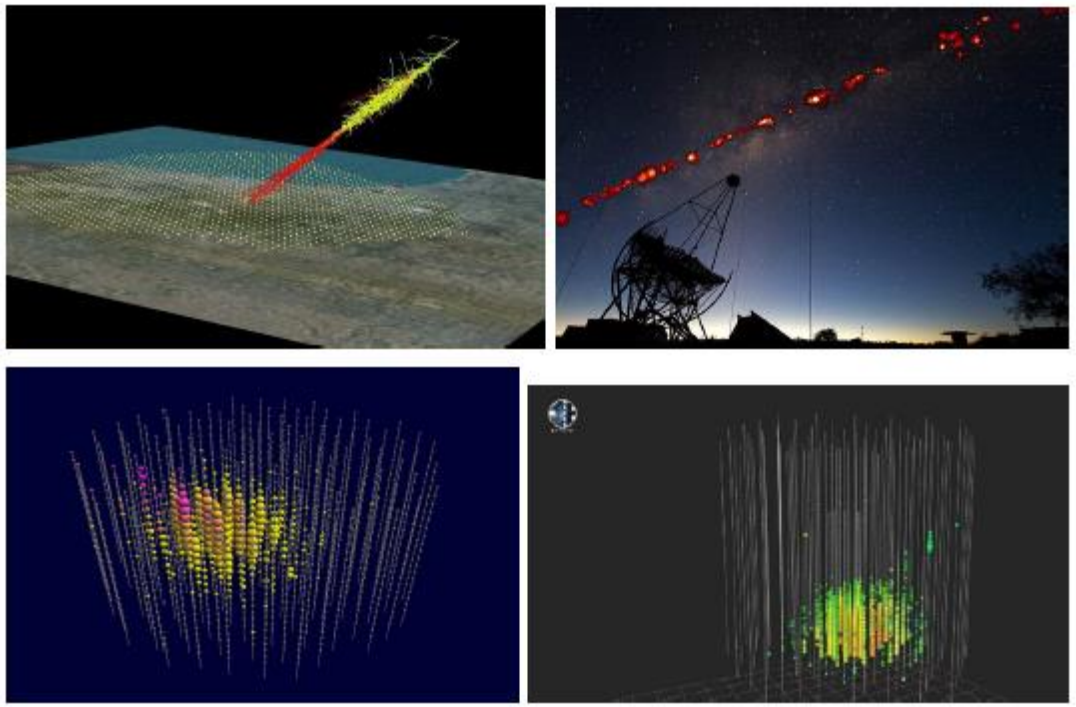
- **Topic A: Development work for the provision of technologies to leverage heterogeneous computing resources**
- **Topic B: Application and testing of virtualized software components in the environment of heterogeneous computing resources**
- **Topic C: Deep learning, gaining knowledge through well-founded data-driven methods**
- **Topic D: Event reconstruction: cost and energy efficient use of computing resources**

**Approved for period 10/2018-9/2021**


one plan of action:
ErUM-Data: Contributions to the digital agenda

# Support by the BMBF (II):

**BMBF call:**
**Funding measure: Digital change in education, science and research**
**Funding area: Kuration criteria and data quality standards**



**Proposal submitted** (5 Institutes; coordinator A. Haungs, KIT)

***FAIR Research Data from High Energy Astroparticles Physics***

- **Preparation of efficient multi-messenger analyses**
- **Preparation for public access to coherent particle shower data**

➔ **Develop concept regarding data and metadata format and quality,**

➔ **Develop concept regarding the curation of combination of the data**

➔ **Including Data and Simulations**

➔ **Develop a demonstrator to validate the concept**

**This also ensures that the data are prepared for free access as basis for a targeted 'Multi-Messenger Open Science' culture in astroparticle physics.**

**for period 1/2019-12/2021**

# Organisational: DPG-AKPIK, Arbeitskreis Physik, moderne Informationstechnologie und Künstliche Intelligenz

**Broad representation of interests for the relevant topics in physics**

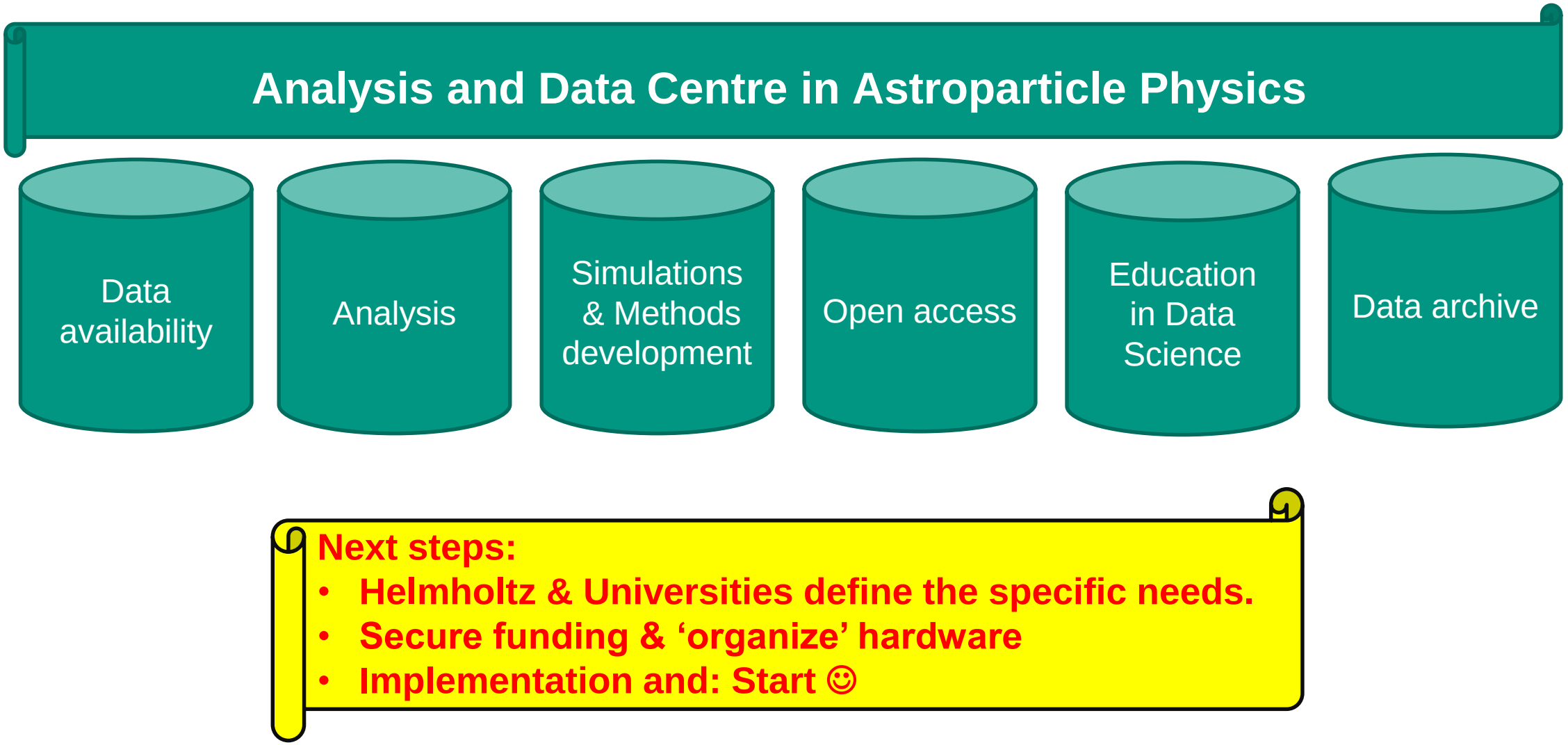

www.dpg-physik.de/dpg/gliederung/ak/akpik/index.html

**Ask Karl Mannheim, Martin Erdmann, Wolfgang Rhode**

**Topics:**

1. **BIG DATA:** archiving, processing, management, analysis and simulation of complex data streams, HPC, information theory, statistical methods
2. **IT:** high-performance data readout systems and mass storage, visualization, smart sensors, bridge technologies for the next level of big data
3. **KI & ROBOTIK:** Data Driven Algorithms & Software, Autonomous Devices, Remote Control, Innovative Applications, Algorithms for Quantum Computers
4. **UNIVERSITY:** curricula and multi-disciplinary research centres, cooperation with the GI Task Force "Data Scientist", IT infrastructure
5. **INDUSTRY and SOCIETY:** Ethics, Technology Assessment, Sustainability, Business, Law, Start-Ups, Public

# 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 ☺**


in Astroparticle Physics

Thanks are going to

- The digital committee of KAT
  (M.Erdmann, M.Bartelmann, M.Schuman, A.Haungs)
- The KAT
- The AKPIK

All involved and contributing to this broad topic