Computing Strategy in the HL-LHC Era:

Astroparticle Physics

Workshop | 6 – 7 May 2020

KAT

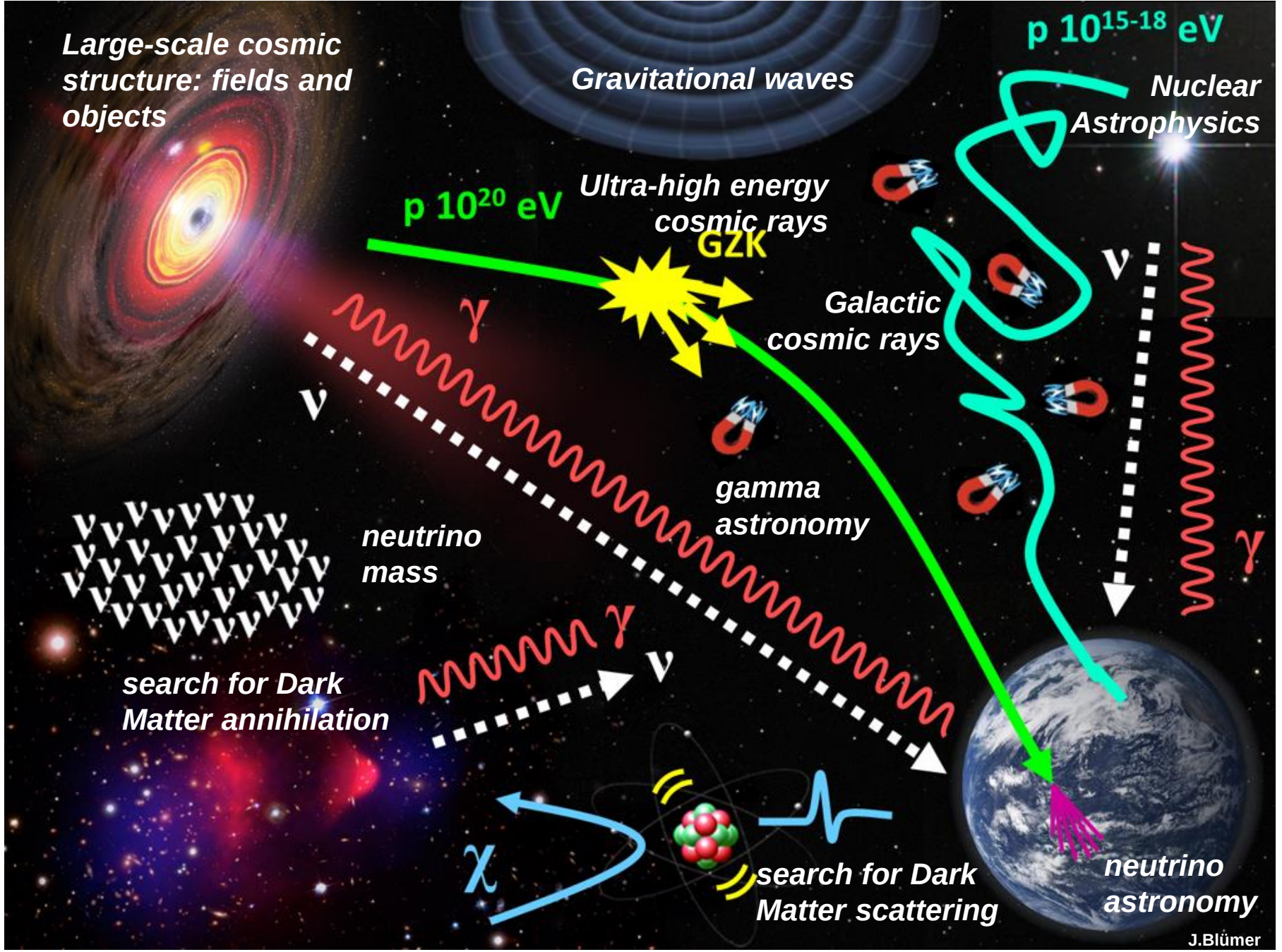
Andreas Haungs, KIT

Astroparticle Physics: Understanding the Multi-Messenger and the Dark Universe

The globally distributed experiments need for an efficient use of the data common

- **Science Data Management**
- **Federated computing infrastructures**

This requires an experiment-overarching agreement



Recommendations of the KAT (white paper)

„Astroparticle Physics in the Light of the Digital Agenda of the Federal Government*“

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

Status Infrastructures in Astroparticle Physics

Computing:

- (Co-use of) Institutional resources (partly WLCG resources)
- GridKa: Tier1-centre in the world wide LHC Computing Grid (e.g. Auger@GridKa)
- Experiment-oriented resources (e.g. CTA@DESY)
- Co-use of facility infrastructures (e.g. IceCube at DESY)
- Moderate use of HPC cluster (Gauß Alliance)

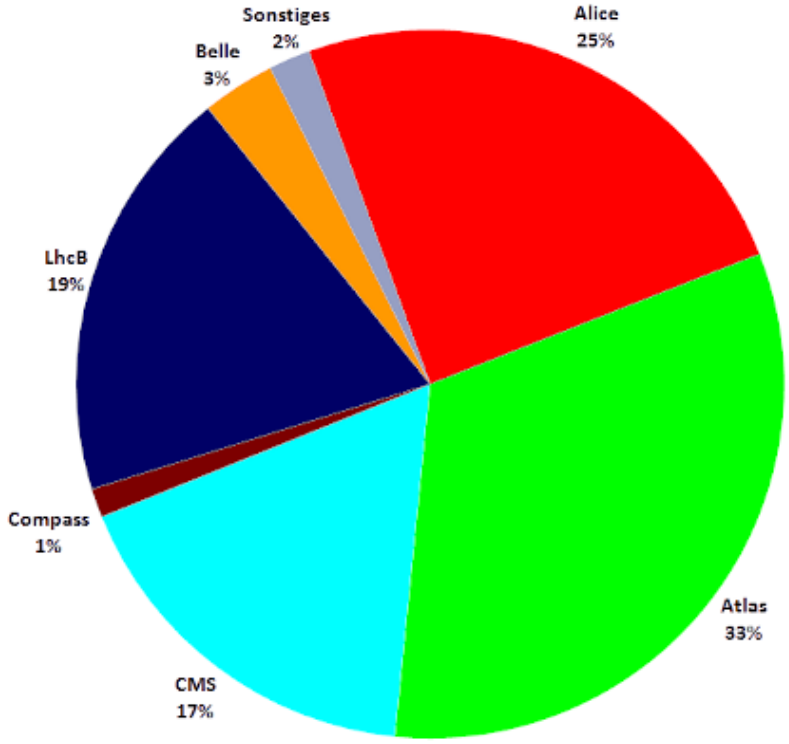
Resarch Data Management:

- KCDC: KASCADE Cosmic ray Data Centre (data access)
- VISPA: to analyze data (Learning Deep Learning)
- GAVO (German Astrophysical Virtual Observatory)
- CERN Open Data Portal (not yet used by APP)

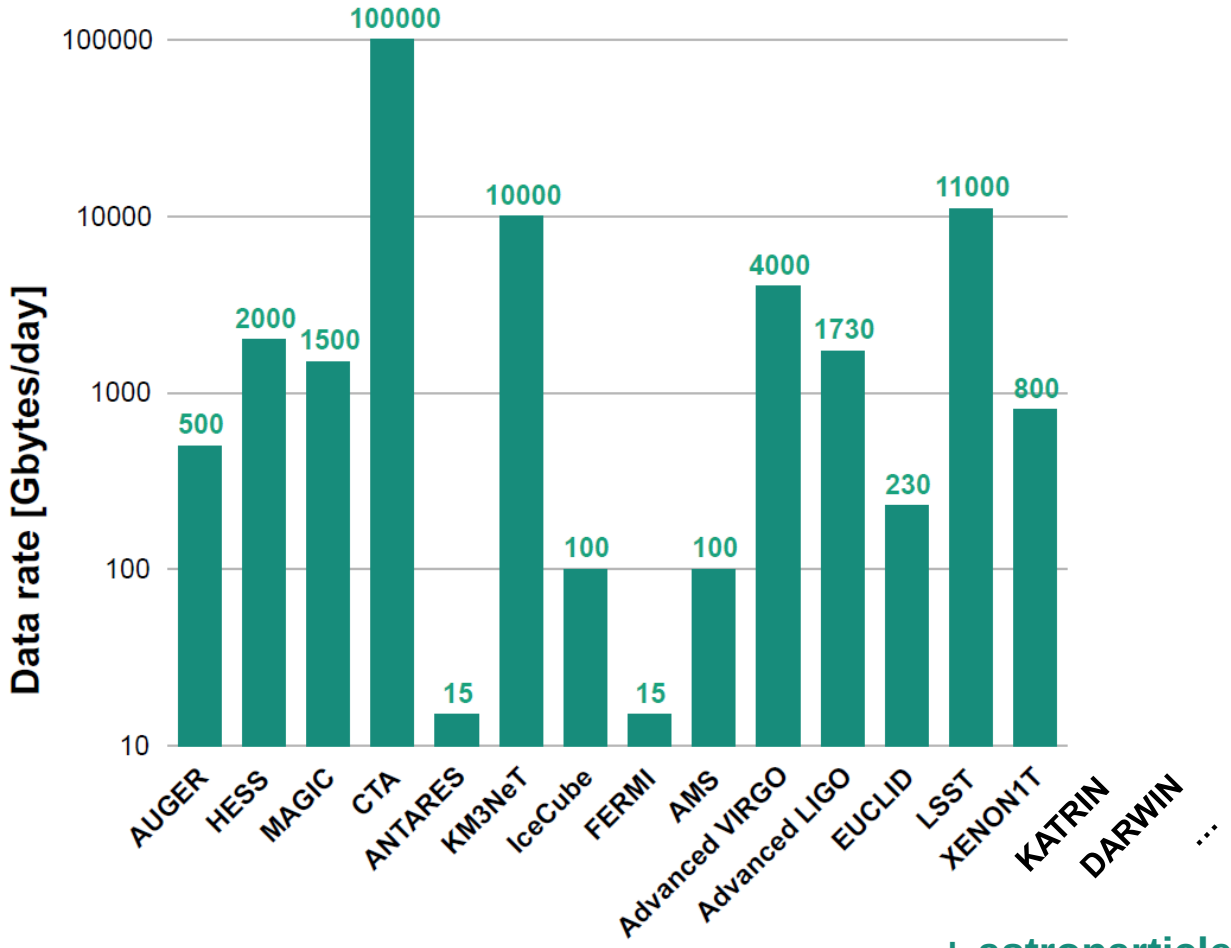
Particle Physics: GridKa (and other Tier-centres)

- 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	28000
Number of compute jobs (last 12 months)	23 million
Number of CPU-hours delivered (last 12 months)	212 million
Disk space	34 PB
Tape space (used)	53 PB



Computing in Astroparticle Physics



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

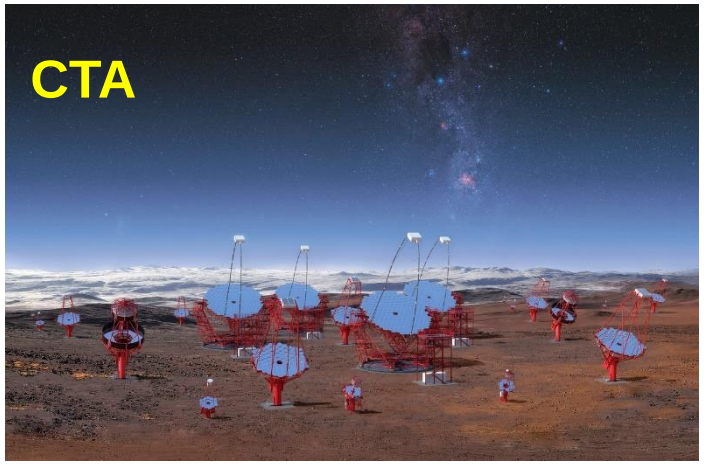
➔Do we need an own Astroparticle Physics computing infrastructure?

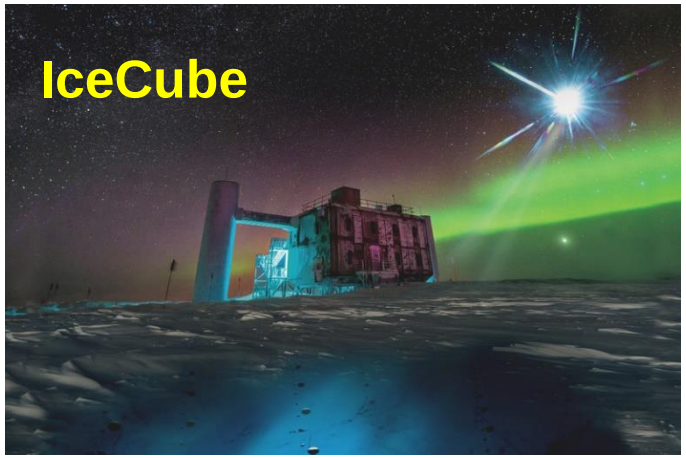
- Synergy with particle physics?
- Grid or Cloud or Lake or other technology?
- Use of commercial providers (amazon, google, …)?
- Is there a relation to NFDI, ErUM-Data, EOSC?

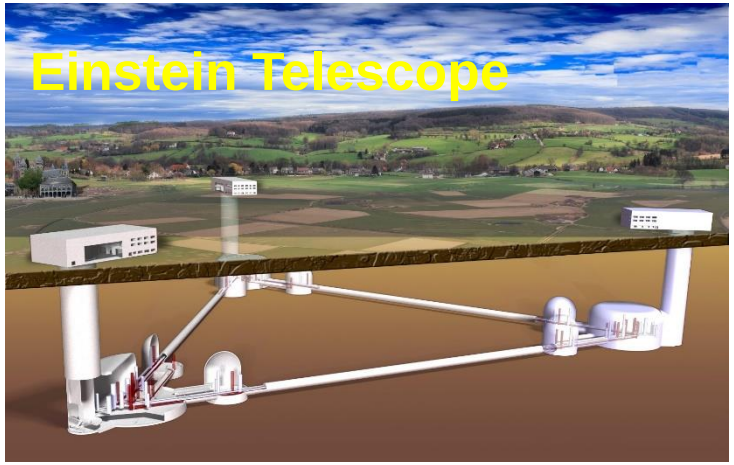
+ astroparticle part of SKA?
+ Einstein Telescope
+ enhanced request from simulations

partly organized in the new faciltes of astroparticle physics

2020+: Flagship Experiments of German Astroparticle Physics (ErUM-Pro)







Example Computing Model:
CTA Science Data Management Centre

cherenkov telescope array

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

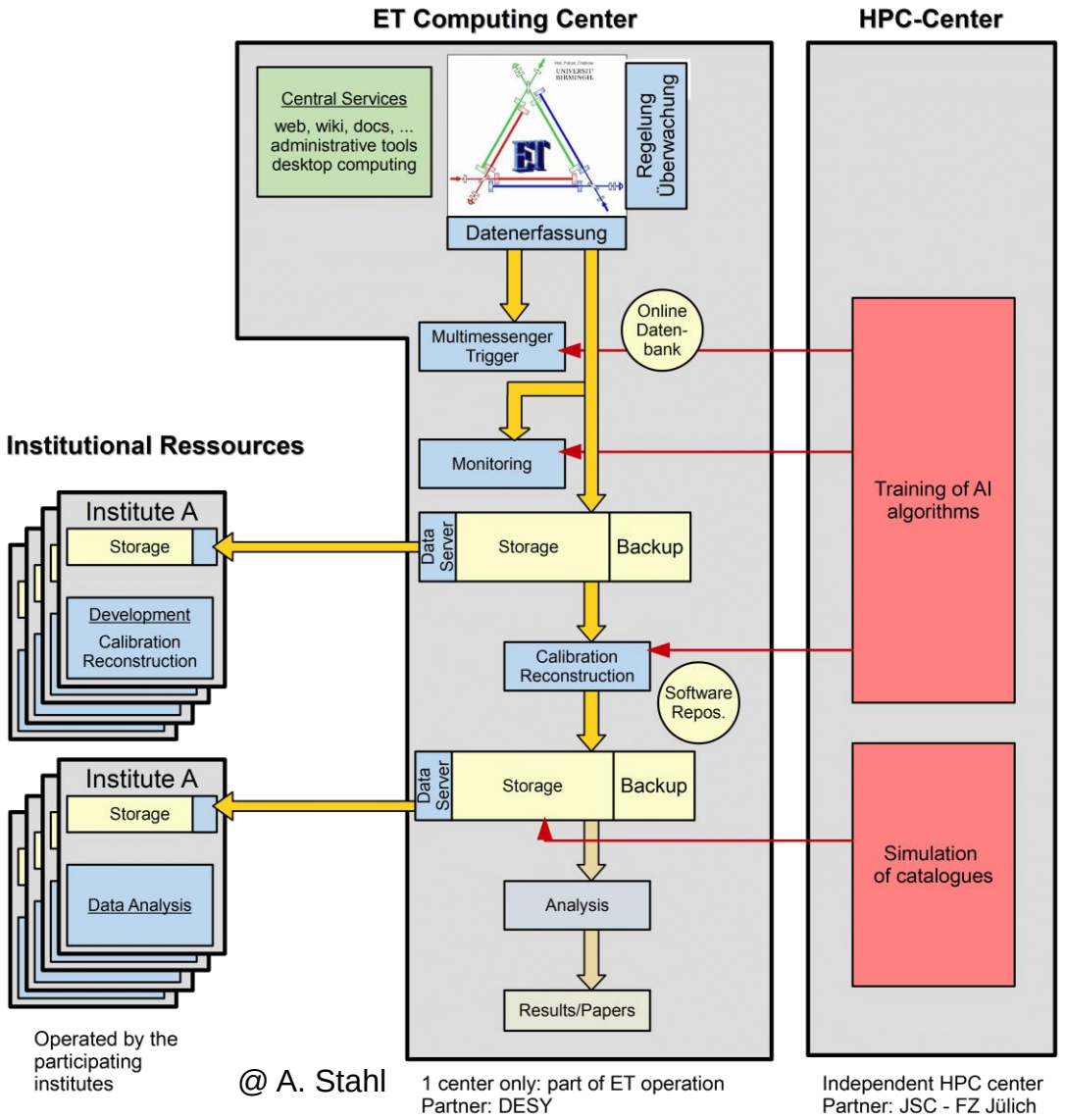
Example Computing Model:
Computing Challenges of Einstein Telescope

Computing Model:

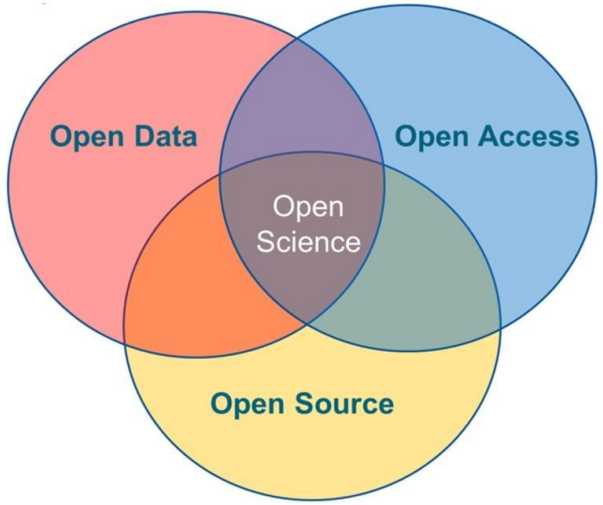
- ET Computing Center, only low latency (= operation costs)
- HPC-Center (= member country costs)
- Institutional Resources (= institutional costs)

Challenge:

- LIGO/Virgo analysis path does not work, since:
  - Many more signals / events
  - Longer signal traces at low frequencies (hours)
  - Parameter set per event much higher (better fit and comparison to template)
  - More parameters available (e.g. polarisation)
  - More types of events, i.e. more template catalogues.
  - Huge amount of (online) monitoring data
- Requests large resources (HPC) for generating and training of catalogues as well as the development of smart algorithms



Goal: Open Science in Astroparticle Physics



- **Open Data – Open Access – Open Source**
  - Provide sustainable access to scientific data
  - Archiving of Data and Meta-Data
  - Providing workflow software and analysis tools
  - Education in Big Data Science
- **Advancement of existing structures at individual facilities (like KCDC and others)**
- **In direction of a virtual Observatory (like in astronomy)**
- **In direction of Tier-systems and DPHEP (like in particle physics)**
- **Consider „Digitale Agenda der Bundesregierung“**
- **OECD Principles and Guidelines for Access to Research Data from Public Funding**
- **Inclusion in data repositories (like re3data.org)**
- **Follow the FAIR principles of data handling (FINDABLE-ACCESSIBLE-INTEROPERABLE-REUSABLE)**

➔ **all this requires a dedicated (federated) infrastructure**

Analysis and Data Center 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:**

Researchers need an environment for simulations and the development of new methods (machine learning).

- **Real-time analysis center:**

The multi-messenger ansatz requires a framework to develop and apply methods for joint data stream analysis.

- **Open access:**

More and more it is necessary to make the scientific data available also to the interested public: public data for public money!

- **Education in data science:**

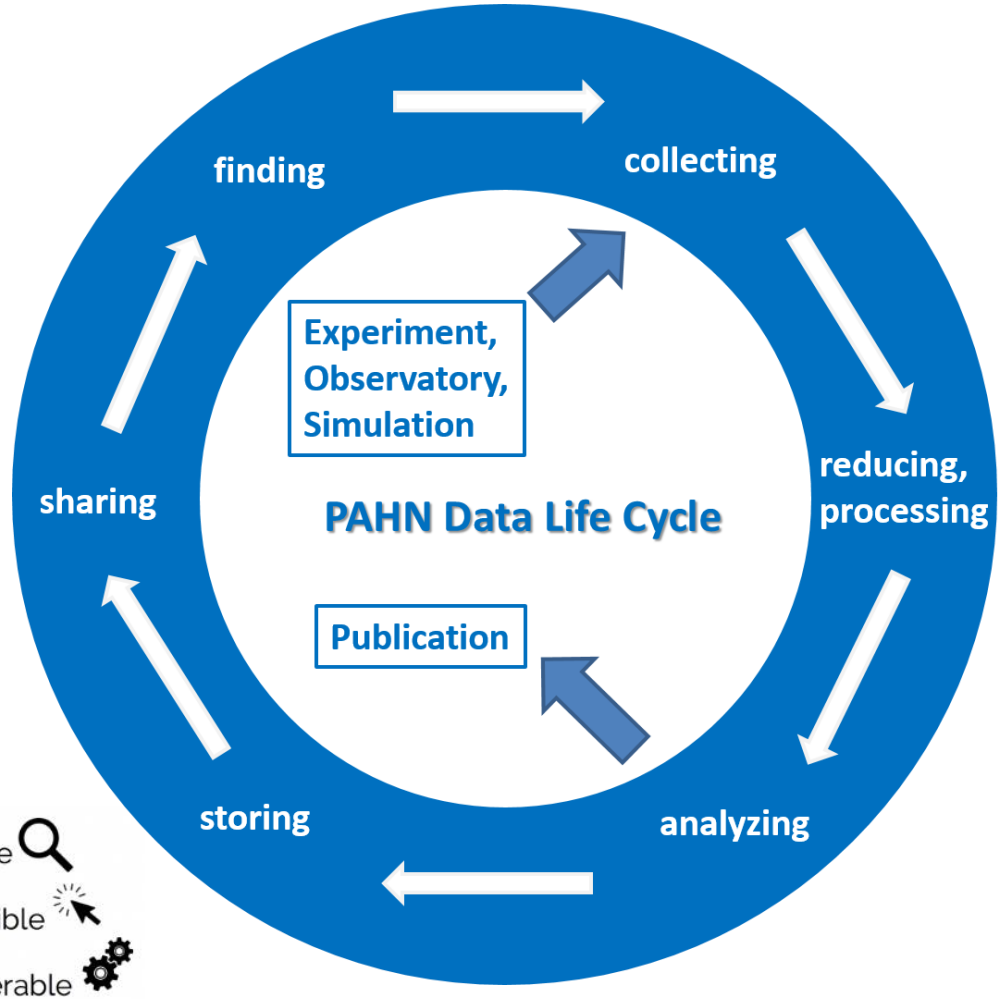
Not only data analysis itself, but also the efficient use of central data and computing infrastructures requires special training.

- **Data archive:**

The valuable scientific data, metadata and software tools must be preserved and remain interpretable for later use (data preservation).

Partly realized in individual experiments or initiatives

FAIR Data Lifecycle Concepts and Open Data



Findable
Accessible
Interoperable
Reusable

Where possible, establish common standards to foster interoperability (between experiments, but also communities)

Importance of "data stewards" as data lifecycle managers and metadata curators (need knowledge from the research field)

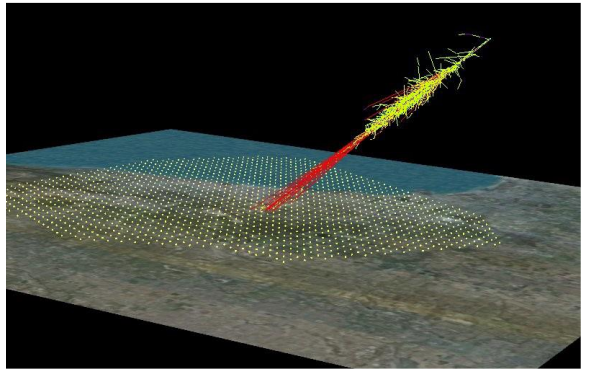
The astroparticle data lifecycle has to provide a FAIR environment for

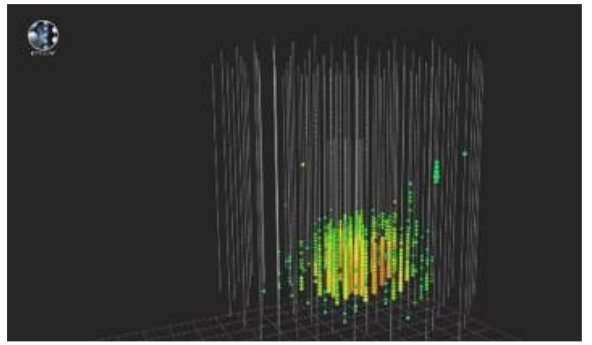
- **(i) data availability**
- **(ii) method development**
- **(iii) data analysis**
- **(iv) big data education**
- **(v) open access**
- **(vi) data archiving**
- **(vii) data mining**
- **(viii) software archiving**

- **Each arrow requires *FAIR* data management**
- **Each step needs appropriate metadata**
- **The cycle includes data, metadata and workflows**

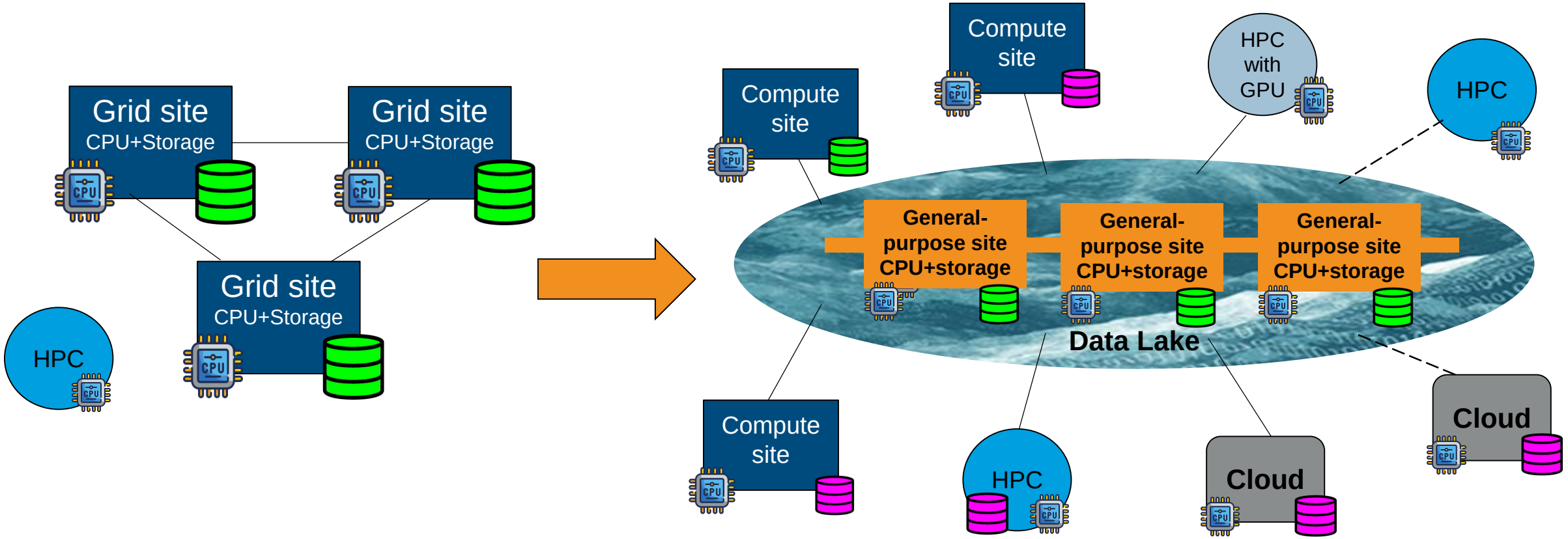
Analysis and Data Centre for Multi-Messenger Astroparticle Physics ADC-MAPP

- **Basics**
  - ADC-MAPP project period 2019-2020
  - funded by Helmholtz
- **Main targets of the Project**
  - Provide sustainable access to scientific data
  - Archiving of Data and Meta-Data
  - Providing analysis tools
  - Foster real-time analysis
  - Education in Big Data Science
  - Development area for multi-messenger analyses (e.g. Deep Learning)
  - Platform for communication and exchange within Astroparticle Physics





Data and Computing Model for the Future



TODAY

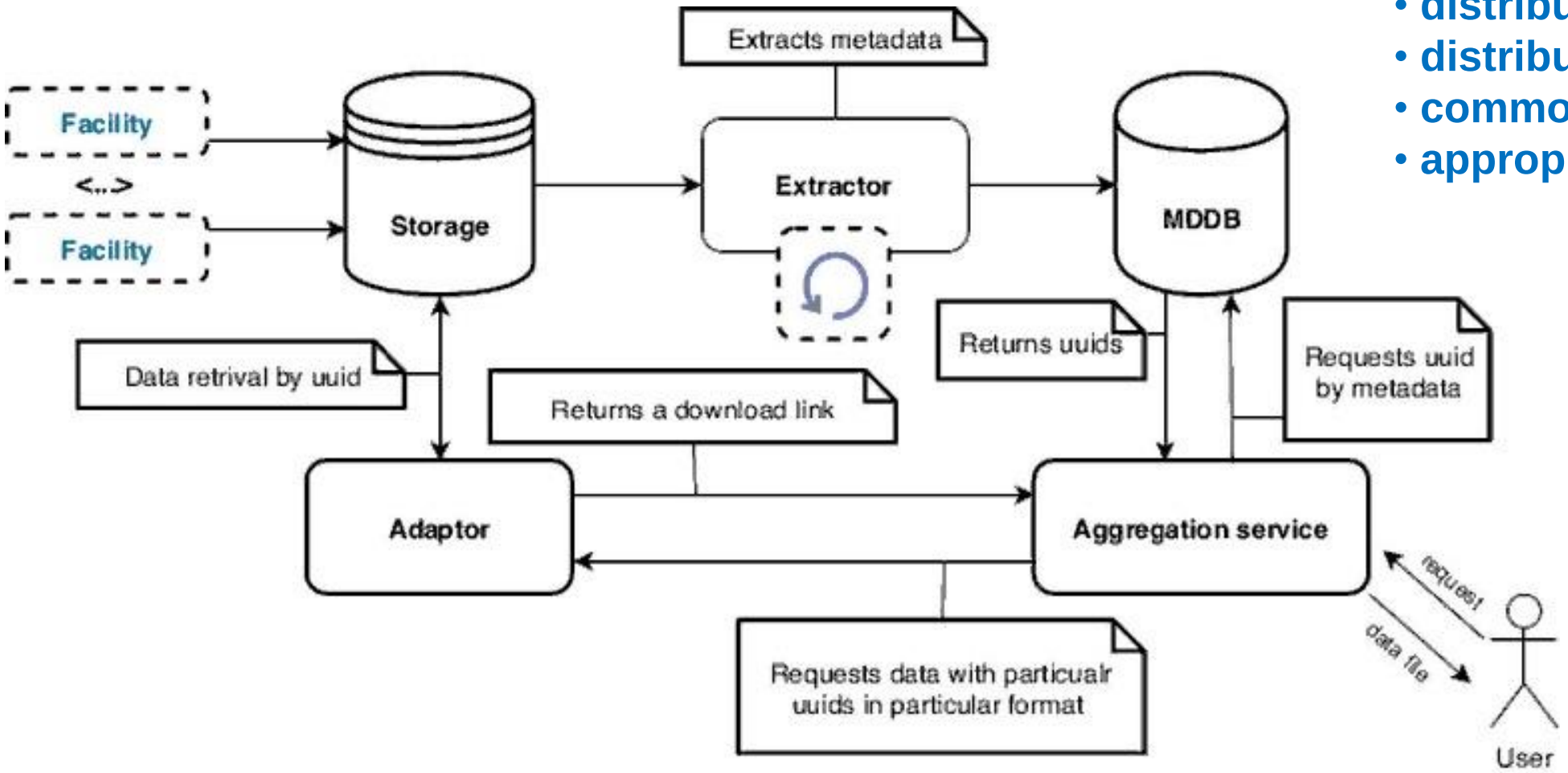
- >170 dedicated grid sites
- Based on high-throughput computing (HTC) architectures
- Connected via dedicated networks
- Data storage at the sites

FUTURE

- Globally distributed data lakes with remote access
- Additional compute resources at clouds and high-performance computing (HPC) centres
- More complex storage architecture (cache)

Astroparticle Data Life Cycle Initiative: Data Aggregation

- distributed data provider
- distributed data storage
- common Metadata definition
- appropriate software



from: KIT-Russia project on Astroparticle DLC (Tokareva)

Federated Infrastructures for Astroparticle Physics (…in Germany)

- **Starting position**
  - **more and more complex experiments and research facilities**
  - **rapidly increasing digitization levels and therefore growing data volumes of the instruments**
  - **sophisticated simulation and data analyses**
  - **request of combination of data from different facilities (Multi-Messenger APP)**

➔ considerably growing needs of the scientific community for an efficient Information and Communication Technology (ICT) infrastructure

- **A scientific (ICT) infrastructure for data-intensive research requires**
  - **large Storage, fast Network, high Computing Power**
- **The future computing model for Astroparticle Physics will have many similarities and synergies with the HEP (HL-LHC) activities**
- **Such an ICT infrastructure for Astroparticle Physics should be seen in the context of broader national research data infrastructures and at the international level, e.g. in the context of European cloud initiatives**
- **Such a common virtual ICT infrastructure should be connected to experiment-specific infrastructures and should foster the inclusion of commercial resources.**

Conclusions:

- **Software development for efficient use of current and future ICT**
  - At federated infrastructures (big data analytics, standard analyses)
  - At local infrastructures (data reduction)
  - For data curation, data quality, data security
  - Use synergies in tool developments (open workflow tools)
  - Sustainable education and training
  - Synergetic developments in FAIR data handling from data source to open data
  - Requires change of culture in recognition of this kind of work

- **Future computing in Astroparticle Physics requires**
  - Participation on the R&D for Exa-Scale machines
  - Access to (new) federated infrastructure (hardware)
  - Concepts of mixture between dedicated and federated infrastructures
  - synergies in using distributed resources
  - user-led computing model
  - Close connections to ErUM-Data, NFDI, ESCAPE, EOSC, …
  - Permanent astroparticle physics dedicated manpower at the infrastructures

➢ **Towards a virtual synergetic (ErUM) computing infrastructure (HTC and HPC) to be co-used by the (German) Astroparticle Physics Community**