# The Digital Agenda of Astroparticle Physics
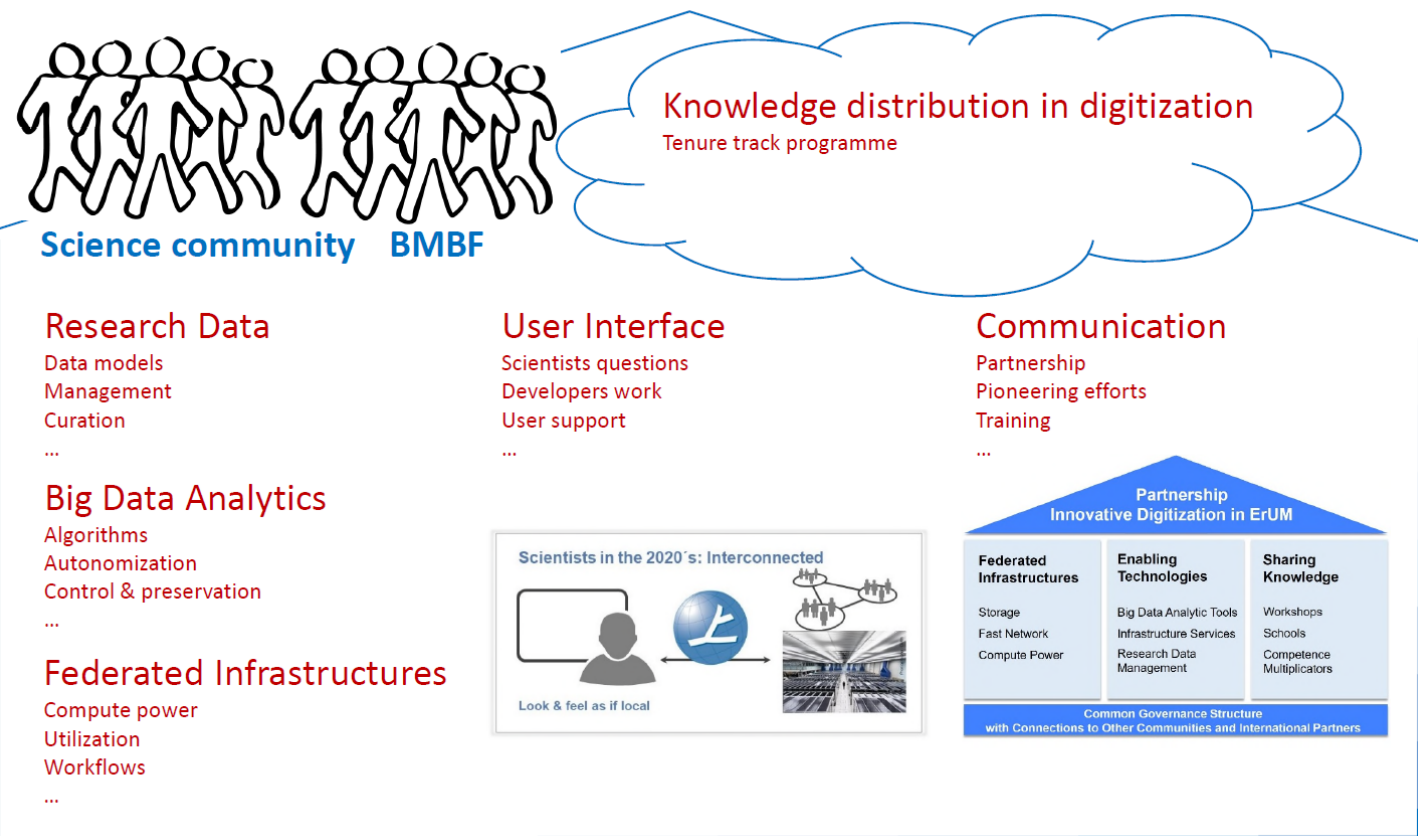
**7th KAT Strategy Meetingmeeting, December 2020**

**Andreas Haungs for the KAT Digital Working Group**

1. **ErUM-Data**

2. **Federated Infrastructures**

3. **PUNCH4NFDI**

Science community    BMBF

Knowledge distribution in digitization
Tenure track programme

**Challenges and Opportunities of Digital Transformation in Fundamental Research on Universe and Matter**

Recommendations of the ErUM Committees
[ ErUM - Exploration of the Universe and Matter ]
29 April 2019

### Research Data
Data models
Management
Curation
...

### Big Data Analytics
Algorithms
Autonomization
Control & preservation
...

### Federated Infrastructures
Compute power
Utilization
Workflows
...

### User Interface
Scientists questions
Developers work
User support
...

Scientists in the 2020´s: Interconnected

Look & feel as if local

### Communication
Partnership
Pioneering efforts
Training
...

**Partnership**
**Innovative Digitization in ErUM**

| Federated Infrastructures | Enabling Technologies | Sharing Knowledge |
|---|---|---|
| Storage | Big Data Analytic Tools | Workshops |
| Fast Network | Infrastructure Services | Schools |
| Compute Power | Research Data Management | Competence Multiplicators |

**Common Governance Structure**
**with Connections to Other Communities and International Partners**

**Martin Erdmann**

Andreas Haungs, KIT-IAP

KAT . Komitee für
Astro . Teilchen . Physik

# 1. ErUM-Data

➢ **"Aktionsplan" ErUM-Data is published since 12/11/2020**

➢ **First action is installation of the ErUM-Data-Hub, a central office for coordination and communication ErUM-Data activities**

➢ **First scientific call, focusing on 'machine learning and artificial intelligence' will be published end of the year.**

➢ **For ErUM-Pro (KET-KHuK) two cross-activity proposals including KAT were prepared:**
   ➢ **Federated Infrastructures (Alexander Schmidt, Aachen, 10 applicants) ➔ submitted 31 October 2020**
   ➢ **Big Data Analytics (Gregor Kasieczka, Hamburg, >20 applicants) ➔ will wait for the ErUM-Data call**

➢ **ErUM-Data community workshop under consideration for mid January**
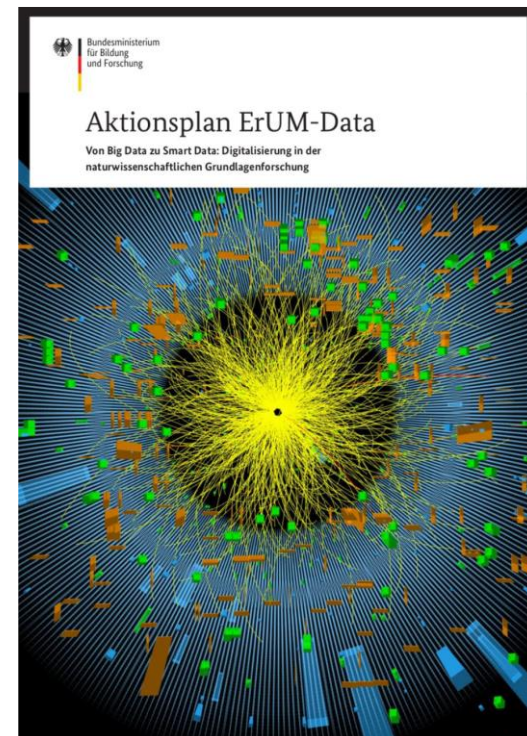


https://www.bmbf.de/upload_filestore/pub/Aktionsplan_ErUM_Data.pdf

Andreas Haungs, KIT-IAP

KAT. Komitee für
Astro.Teilchen.Physik
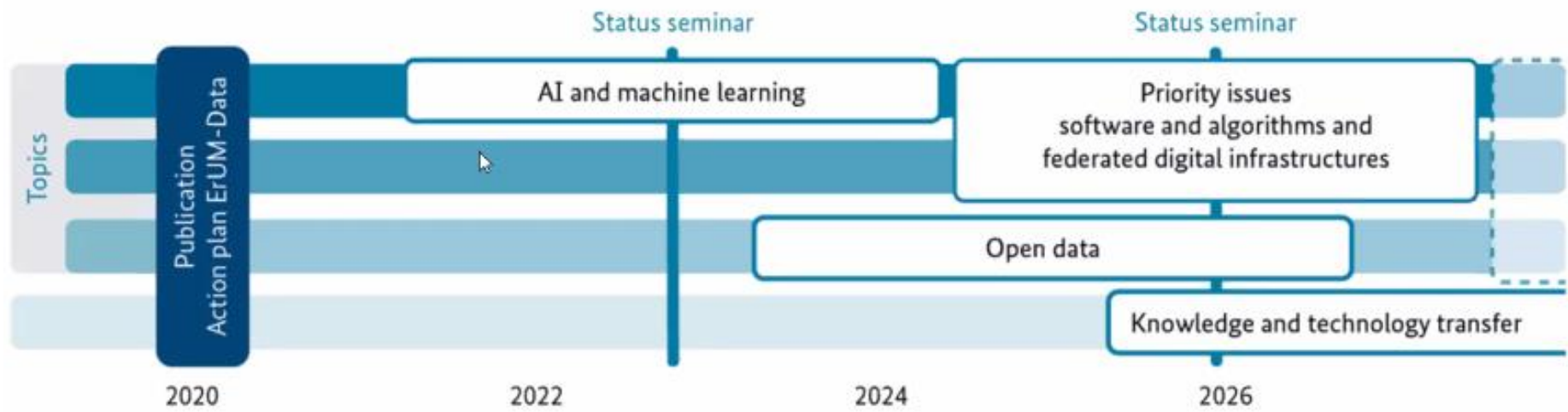
# 1. ErUM-Data



**Action plan provides up to 120 Mio.€**

**E. Lilienthal**

Andreas Haungs, KIT-IAP

KAT. Komitee für
Astro. Teilchen. Physik

# 1. ErUM-Data

## ErUM-Data-Hub

| | Scientists with doctoral degree |
|---|---|
| KFS | 2,300 |
| KFN | 1,600 |
| RDS | 1,500 |
| KHuK | 1,500 |
| KET | 1,300 |
| KAT | 500 |
| KfB | 200 |
| KFSI | 100 |
| | 9,000 |

**Editorial Board (mandate from the DIGUM Overview Board):**
**Erik Bründermann - Martin Erdmann - Christian Gutt - Andreas Haungs – Bridget Murphy - Markus Schumacher - Kilian Schwarz**
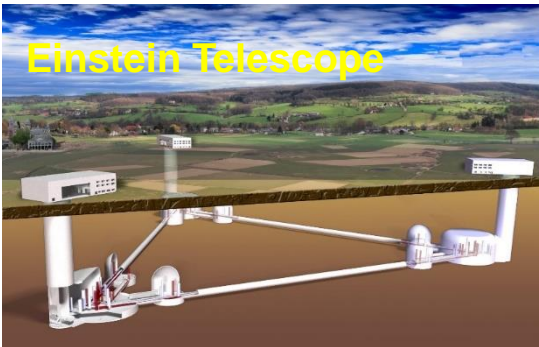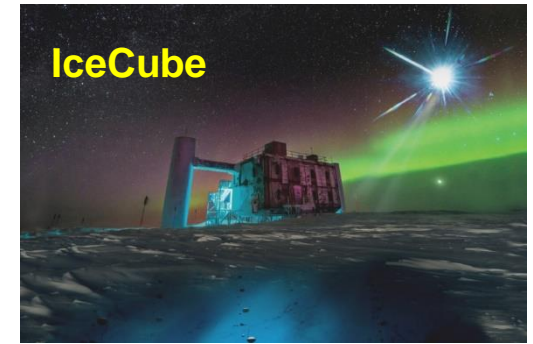
**Organization DIGUM**

**Coordination ErUM-Data HUB**

**Overview Board (OB)**
8 Committee Chairs, 1 Resource Provider, 1 Representative of the BMBF, 1 Representative of DESY-PT

**Speaker / 5 Co-Speakers**

**Digitization Board (DB)**
Speaker, 5 Co-speakers, 8 Experts from committees, 1 Resource Provider

**Resource Provider Board (RB)**
10 Resource Providers, 8 Experts from committees

**Administrative Office (AO)**
Backbone coordination, includes 1 Administration coordinator & Team of assistents and secretaries

**Annual Conference**
of the ErUM-Data Working Groups

**International Advisory Board (IAB)**
ca. 5 from Science, Industry

**Topic Boards**

**Topic Federated Infrastructures**
Board: Coordinators, Experts
Compute power
Utilization
Workflows
...

**Topic Big Data Analytics**
Board: Coordinators, Experts
Algorithms
Autonomization
Control & preservation
...

**Topic Research Data**
Board: Coordinators, Experts
Data models
Management
Curation
...

**Topic User Interface**
Board: Coordinators, Experts
Scientists questions
Developers work
User support
...

**Topic Knowledge distribution**
Board: Coordinators, Experts
Tenure track programme
Workshop, schools
...

Currently community organized

Apply for BMBF HUB-funding

6

KAT. Komitee für Astro. Teilchen. Physik

# 2. Federated Infrastructures

- **May 2020: KET-KHuK-KAT-RDS organized workshop: Computing Strategy in the HL-LHC Era**
  - **Input for the KET-KHuK ErUM-Pro Strategy Meeting**

- **A main result of the workshop discussions was that in this process the high-energy physics community will open the strategy and the resources to other communities (proportional in output and input and for mutual benefit) where naturally KAT and KHuK are the closest partners.**

- **In this process a survey was started to have a guess on more or less reliable numbers from the Astroparticle Physics community.**

Andreas Haungs, KIT-IAP

KAT . Komitee für
Astro . Teilchen . Physik

# 2. Federated Infrastructures

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



Pierre Auger Observatory



CTA



IceCube



Einstein Telescope



KATRIN



GERDA/LEGEND



DARWIN

Andreas Haungs, KIT-IAP

KAT . Komitee für
Astro . Teilchen . Physik

# 2. Federated Infrastructures

**Example Computing Model:**
**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

Andreas Haungs, KIT-IAP

# 2. Federated Infrastructures

## 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**
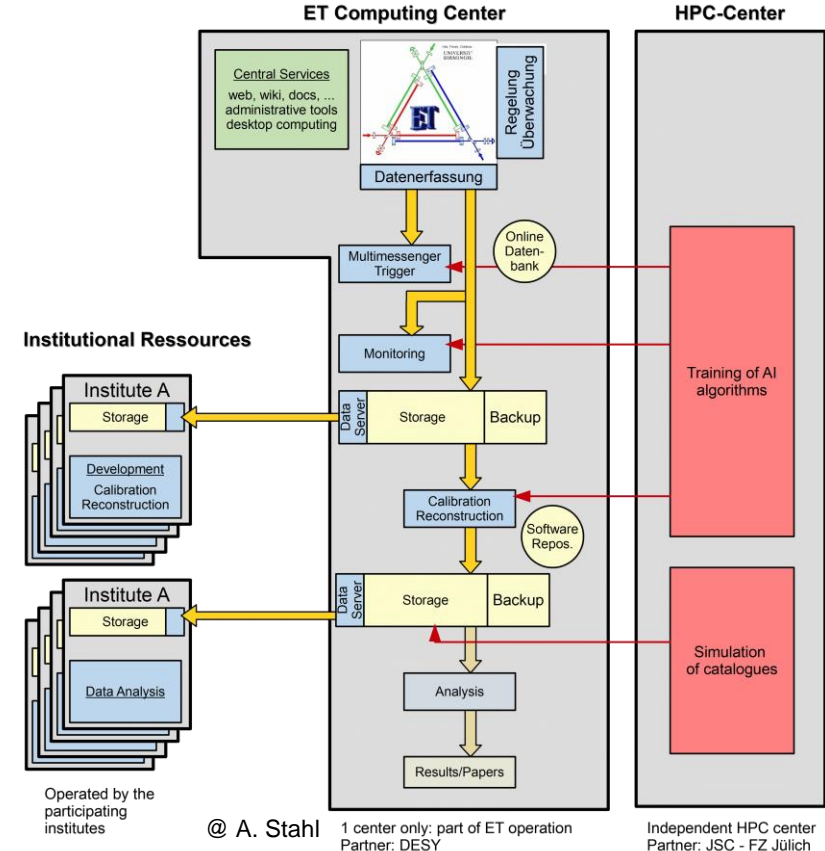


@ A. Stahl

1 center only: part of ET operation
Partner: DESY

Independent HPC center
Partner: JSC - FZ Jülich

Andreas Haungs, KIT-IAP

# 2. Federated Infrastructures

**Assessment of the demand for federated resources in computing of APP:**

- **Total German share of computing requests of the ErUM-Pro projects in addition to usage of institutional resources.**
- **\* CTA: 2/3 are provided by the CTAO**
- **+ Theorie: use of supercomputers of major part, e.g. in Jülich, SuperMUC,……**
- **# IceCube: ca. 70% currently provided by the Tier1 in DESY-Zeuthen (in 2028 ca. 30%)**

- **Projected requests for 2028: factor ~4 for CPU, Disk and Tape, factor ~10 for GPU (mainly by ET)**

| Request in 2021 | Auger | IceCube | CTA* | ET | KATRIN | GERDA / LEGEND | DARWIN | Multi-Messenger | Theorie | Summe |
|---|---|---|---|---|---|---|---|---|---|---|
| CPU [CPU-years] | 500 | 2000 | 1500 | 0 | 500 | n/a | 0 | 100 | 2000 | 6600 |
| GPU [GPU-years] | 40 | 400 | 0 | 0 | 0 | n/a | 0 | 50 | 100 | 590 |
| Disk [PB] | 0.8 | 2 | 1.6 | 0 | n/a | n/a | 0 | 0.2 | 0.1 | 4.7 |
| Tape [PB] | 3 | 10 | 0 | 0 | n/a | n/a | 0 | 0 | 0 | 13 |

| Projected for 2028 | Auger | IceCube | CTA | ET | KATRIN | GERDA / LEGEND | DARWIN | Multi-Messenger | Theorie | Summe |
|---|---|---|---|---|---|---|---|---|---|---|
| CPU [CPU-years] | 800 | 3000 | 3000 | 5000 | 600 | n/a | 2500 | 1000 | 3000 | 18900 |
| GPU [GPU-years] | 70 | 600 | 0 | 5000 | 400 | n/a | 13 | 500 | 500 | 7083 |
| Disk [PB] | 1.5 | 2 | 7 | 2 | n/a | n/a | 1.9 | 2 | 0.2 | 16.6 |
| Tape [PB] | 5 | 20 | 30 | 0 | n/a | n/a | 1.1 | 4 | 0 | 60 |

Andreas Haungs, KIT-IAP

KAT. Komitee für
Astro . Teilchen . Physik

# 2. Federated Infrastructures

**Assessment of the demand for federated resources in computing of APP:**

- **To WLCG system projected requests of German share of computing requests of the ErUM-Pro projects in addition to usage of institutional resources.**
- **Projected requests for 2028: factor ~8 for CPU, ~5 Disk and ~10 Tape, factor ~20 for GPU (mainly due to ET)**

- **Theory: The current needs are met by federal or state-operated supercomputer centers (Jülich SC, Leibniz Center Munich, HLRN, etc.); not clear if this is possible for 2028.**

| Request in 2021 | Auger | IceCube | CTA* | ET | KATRIN | GERDA / LEGEND | DARWIN | Multi-Messenger | Theorie | Summe |
|---|---|---|---|---|---|---|---|---|---|---|
| CPU [CPU-years] | 500 | 500 | 500 | 0 | 500 | n/a | 0 | 100 | 0 | 2100 |
| GPU [GPU-years] | 40 | 200 | 0 | 0 | 0 | n/a | 0 | 50 | 0 | 290 |
| Disk [PB] | 0.8 | 1 | 0.5 | 0 | n/a | n/a | 0 | 0.2 | 0 | 2.5 |
| Tape [PB] | 3 | 0 | 0 | 0 | n/a | n/a | 0 | 0 | 0 | 3 |

| Projected for 2028 | Auger | IceCube | CTA | ET | KATRIN | GERDA / LEGEND | DARWIN | Multi-Messenger | Theorie | Summe |
|---|---|---|---|---|---|---|---|---|---|---|
| CPU [CPU-years] | 800 | 2000 | 1000 | 5000 | 600 | n/a | 2500 | 1000 | 1000 | 13900 |
| GPU [GPU-years] | 70 | 400 | 0 | 5000 | 400 | n/a | 13 | 500 | 300 | 6670 |
| Disk [PB] | 1.5 | 2 | 3 | 2 | n/a | n/a | 1.9 | 2 | 0.2 | 12.6 |
| Tape [PB] | 5 | 10 | 10 | 0 | n/a | n/a | 1.1 | 4 | 0 | 30 |

Andreas Haungs, KIT-IAP

KAT. Komitee für Astro·Teilchen·Physik

# 2. Federated Infrastructures

**Text as result of this assessment:**

The demand for computing resources for astroparticle physics in Germany will increase considerably in the coming years. In 2020, the computing for the German flagship experiments (Auger, CTA, IceCube, ET, KATRIN, Gerda/Legend, DARWIN, Multi-Messenger, Theory) will mainly be carried out via institutional, experiment-specific or, as in the case of theory, federated supercomputer resources and only to a small extent via the German WLCG network. An estimation of the 2021 requirements for the German fair-share of the computing of the international experiments resulted in a sum of 2,000 CPU years, 300 GPU years, 2.5 PB disk space and 3 TB tape capacity, which are already largely covered by the WLCG (Tier-1 and Tier-2). A projection into the year 2028 showed an increased demand of about factor 8 in CPU years, factor 20 in GPU years, factor 5 in disk space and factor 10 in tape capacity.

Andreas Haungs, KIT-IAP

KAT. Komitee für
Astro. Teilchen. Physik

# 3. PUNCH4NFDI

**PUNCH4NFDI represents particle physics, astrophysics, astroparticle physics and hadron & nuclear physics.**

- Formed from the consortia ASTRO@NFDI and PAHN-PaN.

- Together more than 4500 scientists received their doctorates.

- Support of the representations KET, KAT, KHuK, RdS.

**ASTRO@NFDI and PAHN-PaN were not recommended for funding in the first round of NFDI, but**

- both requested to resubmit.

- Reviewers name clear strengths and unique selling points.

**DFG-Fachkollegien 309 „Teilchen, Kerne und Felder"
sowie 311 „Astrophysik und Astronomie".**

➔ **Evaluation second round: 10/12/2020 (online)**

The **PAHN-PaN** Consortium
Particle, Astroparticle, Hadron & Nuclear Physics accelerate the NFDI

PAHN-PaN

For the PAHN-PaN Consortium:
Andreas Haungs (KIT), Gregor Kasieczka (Hamburg),
Arnulf Quadt (Göttingen), Thomas Schörner (DESY),
Kilian Schwarz (GSI)

Bonn-Königswinter, 5 December 2019

DFG Deutsche Forschungsgemeinschaft

Astro@NFDI

Astrophysics and Astro Particle Physics
in the National Research Data Infrastructure

| Matthias Steinmetz | Harry Enke | Michael Kramer | Susanne Pfalzner | Stefan Wagner |
| AIP, PI | AIP, PM | MPIfR, Co-I | FZJ, Co-I | ZAH, Co-I |

PUNCH 4NFDI

KAT . Komitee für
Astro . Teilchen . Physik

Particle physics
(Higgs event in CMS)
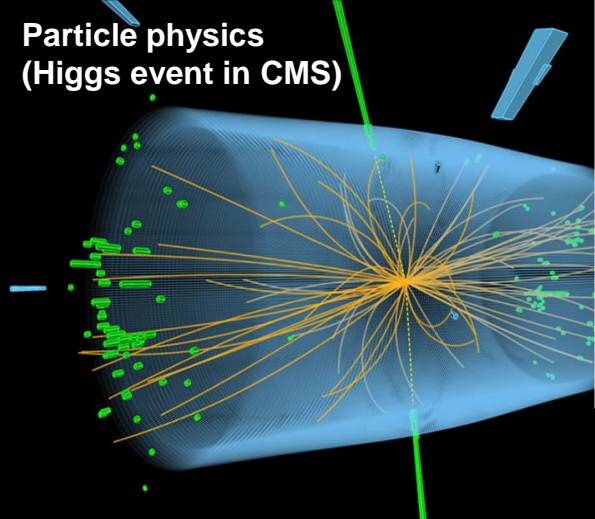
ALPS-II@DESY
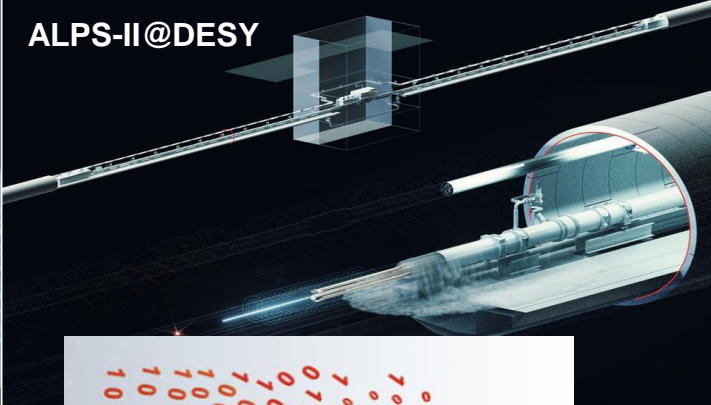
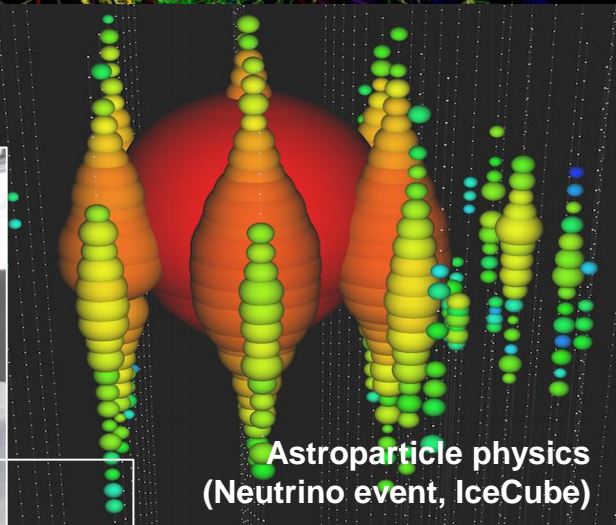Hadron&Nuclear physics
(Heavy-ion collision, Alice)
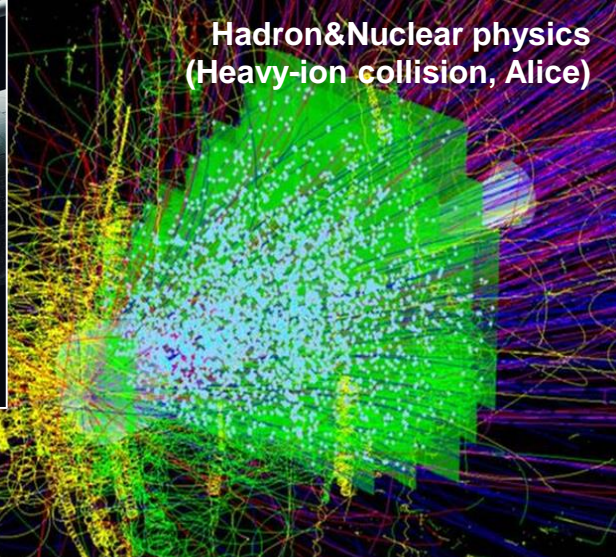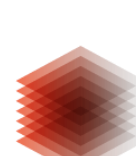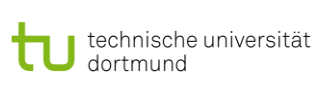
PUNCH
4NFDI

Astronomy
(Black hole, Event Horizon Telescope)
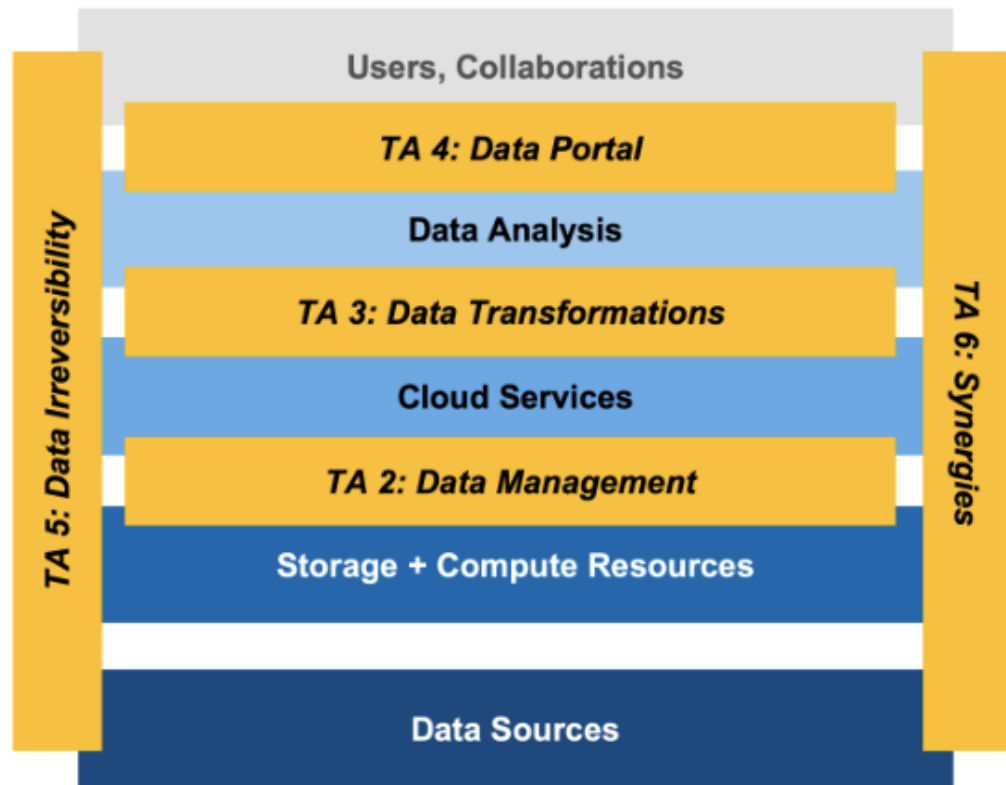
JUWEL@FZJ

Astroparticle physics
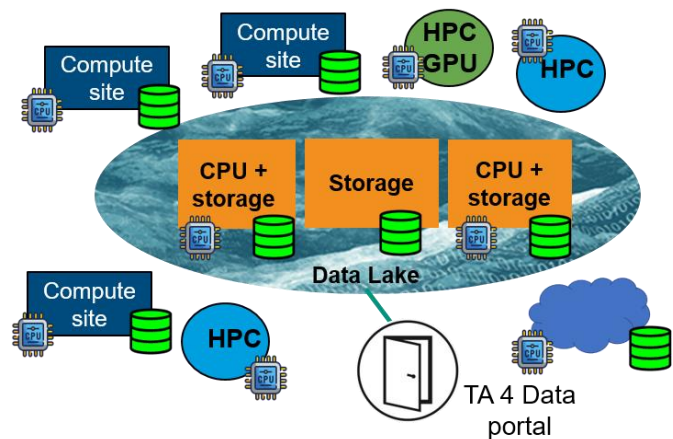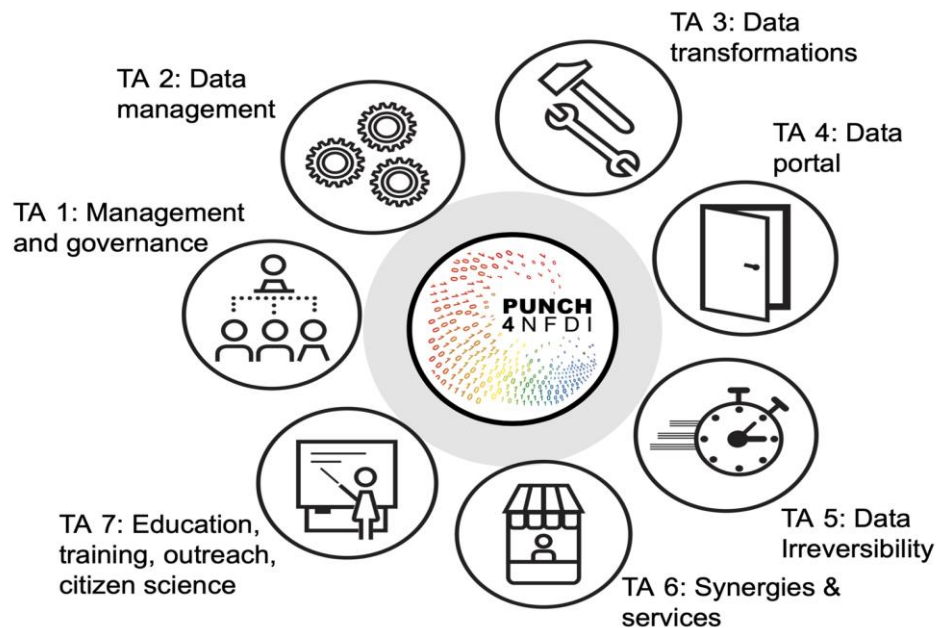(Neutrino event, IceCube)

Andreas Haungs, KIT-IAP

**Future computing and FAIR research data mamangement in PUNCH is based on a layer model**

Andreas Haungs, KIT-IAP

KAT. Komitee für
Astro. Teilchen. Physik

# 3. PUNCH4NFDI

## Task areas



TA 2: Data management

TA 3: Data transformations

TA 4: Data portal

TA 1: Management and governance

PUNCH4NFDI

TA 5: Data Irreversibility

TA 7: Education, training, outreach, citizen science

TA 6: Synergies & services

## Services



Science data portal

Data access & management services

Computing & storage resource services

PUNCH4NFDI services

Metadata services

Analysis software & data irreversibility services

Andreas Haungs, KIT-IAP

KAT. Komitee für Astro. Teilchen. Physik

# Users to be served



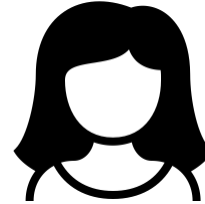| Reviewer | Comparer | Modeller | Combiner | Learner | Interactor |
|---|---|---|---|---|---|
| Would like to know workflow, implementation, data, simulation and statistical analysis and interpretation of the work he is refereeing | Would like compare her results with other findings to reach a deeper understanding by adding information from elsewhere | Would like to maximise result dissemination from his simulations by sharing his data and code with other researchers | Would like to combine her observational/ experimental results with those obtained by others with other instruments | Would like to learn the science and the data by immersing into the real world of science | Would like to obtain first results directly while the observation/ experiment is running to be able to influence it directly |
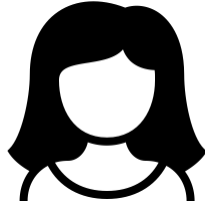
Andreas Haungs, KIT-IAP

KAT. Komitee für Astro. Teilchen. Physik

# 3. PUNCH4NFDI

**PUNCH4NFDI**

- is a consortium of 4 internationally operating communities
- has existing infrastructures as large-scale data providers
- uses customized data management for thousands of users
- challenges Exabyte solutions in terms of data volumes and rates

**PUNCH4NFDI will**

- setup a comprehensive science data platform
- provide a NFDI-wide marketplace for tools, methods, and services
- serve a wide range of use-classes and users
- incubate new technologies in Big Data handling
- train the next generation of scientists and the public in Big Data
- make scientific data rigorously public in a FAIR and sustainable way

**PUNCH4NFDI can**

- assist the NFDI to tackle medium- and long-term challenges in data handling and management,
- keep (the German science system) German physics competitive in international research data management
- strengthen the (industry location) society via technology development and training

Twitter: @PUNCH4NFDI
Mail: punch4nfdi@desy.de
Web: www.punch4nfdi.de

Andreas Haungs, KIT-IAP

KAT . Komitee für
Astro . Teilchen . Physik

# The digital agenda of astroparticle physics in Germany…

**…shows many activities in addition to the three discussed today, e.g.**

- AKPIK
- ESCAPE
- Helmholtz initiative ADC-MAPP
- Helmholtz incubators (AI, Metadata…)
- …..