

KAT Strategy Paper: Overarching Topics

Computing and Data

....an intro for discussion!



KAT@KIT 16-18 October 2024

Andreas Haungs, Institute for Astroparticle Physics

www.kit.edu

(1) Computing and Data: Status in APP

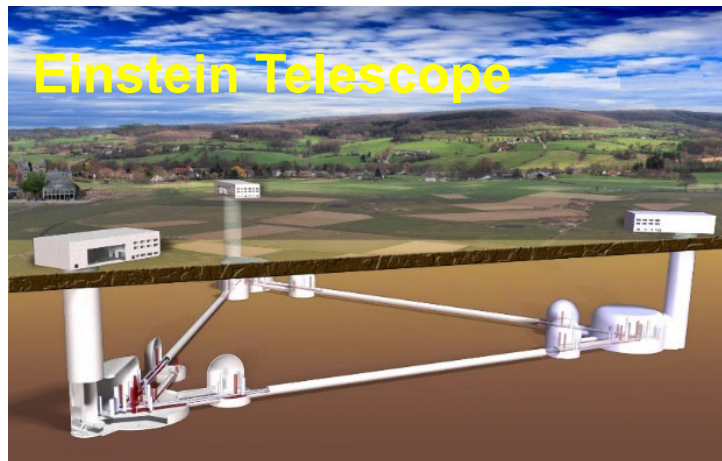
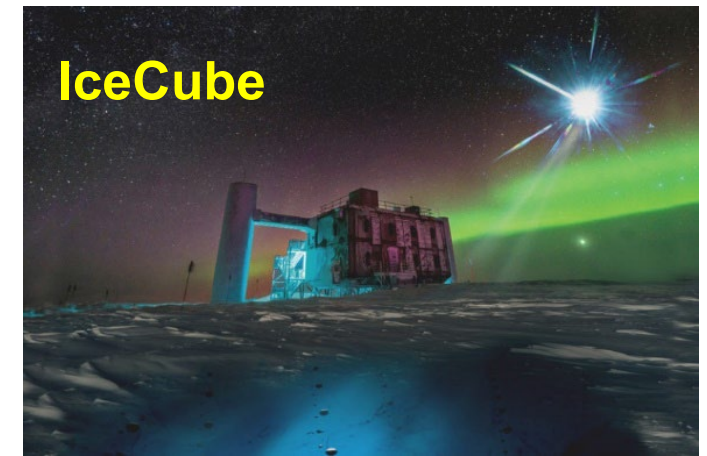


Computing in Astroparticle Physics

- **May 2020: KET-KHuK-KAT-RDS organized workshop: Computing Strategy in the HL-LHC Era**
- **In this process KAT started a survey (2020) to have a guess on more or less reliable numbers now and projected to the HL-LHC era from the German Astroparticle Physics community.**
 - ➔ **due to covid not much changed in the results of this survey....is this true?**
- **However, meanwhile:**
 - **new NHR centers in Germany with new structures and demands**
 - **ErUM-Data calls**
 - **PUNCH4NFDI is active**
 - **KET, in particular, is pushing changes in computing models for HL-LHC**
- **Is a (Re-)Initiation of the discussion on computing strategies and use of federated (computing) infrastructures for Astroparticle Physics in Germany necessary ?!**

Flagship Experiments of German Astroparticle Physics

2020+: Consortia in ErUM-Pro



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

Example Computing Model

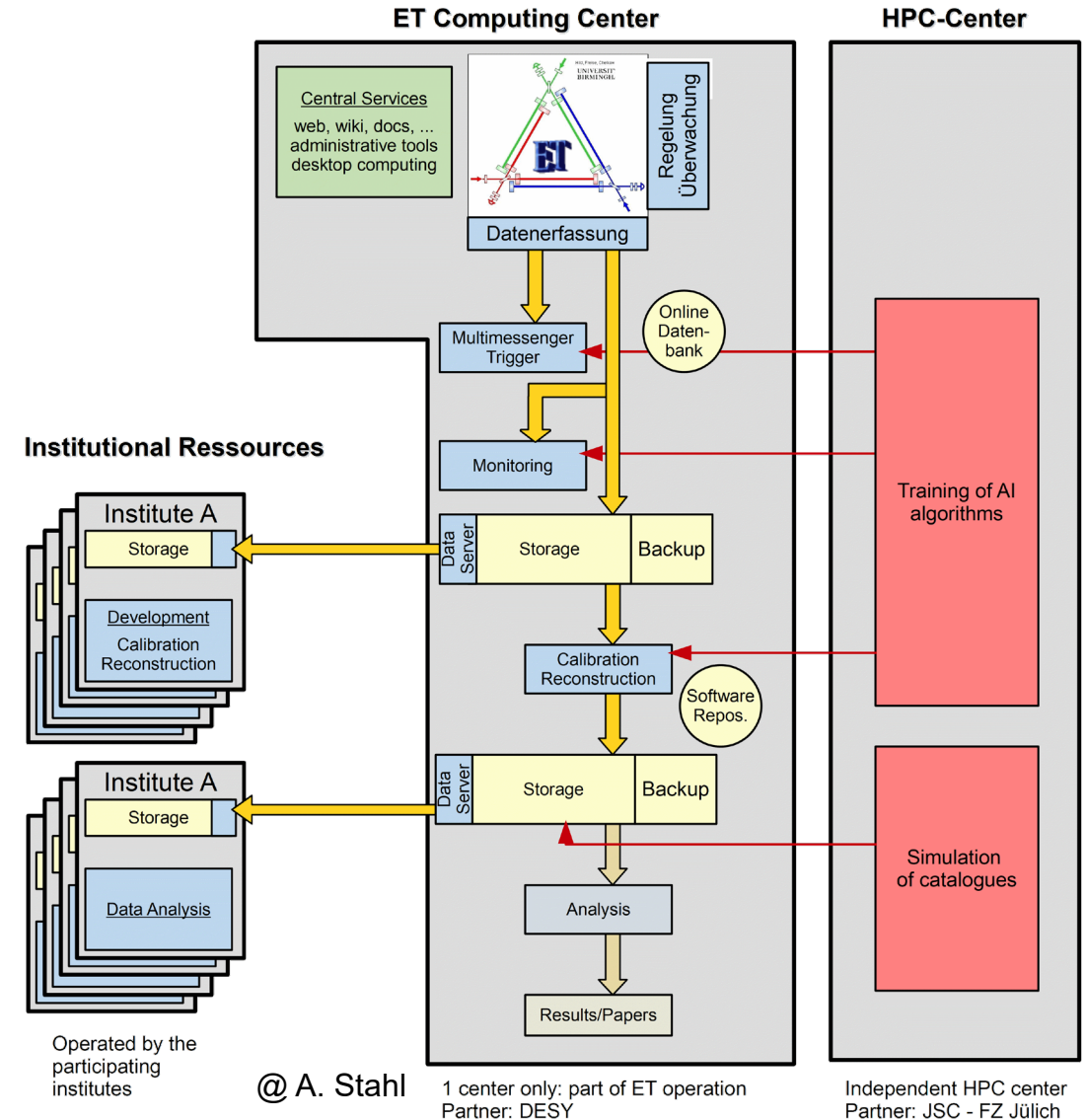
Computing Challenges of Einstein Telescope

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



Survey Results for German Astroparticle Physics (2021)

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

Survey Results for Federated Computing Infrastructures (2021)

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

KAT statement: Federated Infrastructures (2021)

Text as result of this assessment:

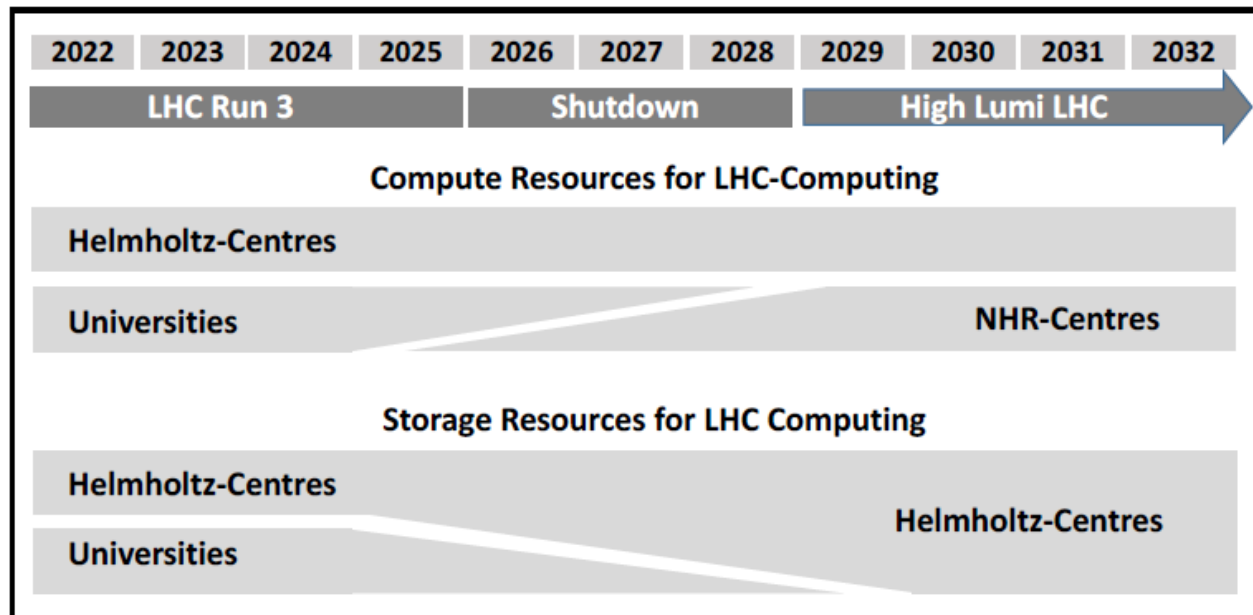
The demand for computing resources for astroparticle physics in Germany will increase considerably in the coming years. In 2021, the computing for the German flagship experiments (Auger, CTA, IceCube, ET, KATRIN, Gerda/Legend, DARWIN, Multi-Messenger, Theory) will mainly be carried out via institutional, 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.

KET Computing Strategy for HL-LHC

Transformed Model for WLCG Resources in Germany

Strategic vision by particle physics community in Germany for HL-LHC era:

- Complete pledged WLCG mass storage to be provided by Helmholtz Centres (DESY, GSI, KIT) and MPP
- CPU pledge from University Tier-2 Centres to be provided by NHR centres
strong believe that NHR centres predestined for this task according to their mandate
(consensus after long and wide discussion btw. all involved parties and current resource providers)



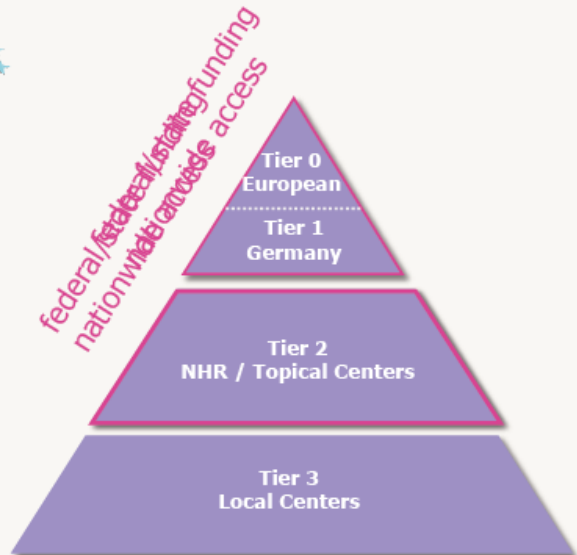
Transition period 2025-2028:

- already available hardware will be operated at university sites
- new resources to be installed and provided at Helmholtz and NHR centres

NHR Computational Physics Symposium (7+8/11/2022);
<https://events.uni-paderborn.de/event/403/>

Markus Schumacher for KET, 8 Nov 2022

NHR as Part of German HPC Infrastructure



- Recommendation by Research Council to introduce Tier-2 National High Performance Computing (NHR) infrastructure
- Competitive applications in 2020
 - official start: Jan 1, 2021
 - total funding 625M Euro (2021-2030)
 - currently 9 NHR centers
- Key aspects
 - joint federal/state-funding
 - transition from regional to competence-oriented for nationwide use
 - free access for all researchers from German universities
 - strengthen methodological competences through coordinated training, continuing education of users
 - specific support for young scientists

November 7, 2022

NHR Computational Physics Symposium

**Christian Plessl,
Paderborn**



NHR Computational Physics Symposium (7+8/11/2022);
<https://events.uni-paderborn.de/event/403/>

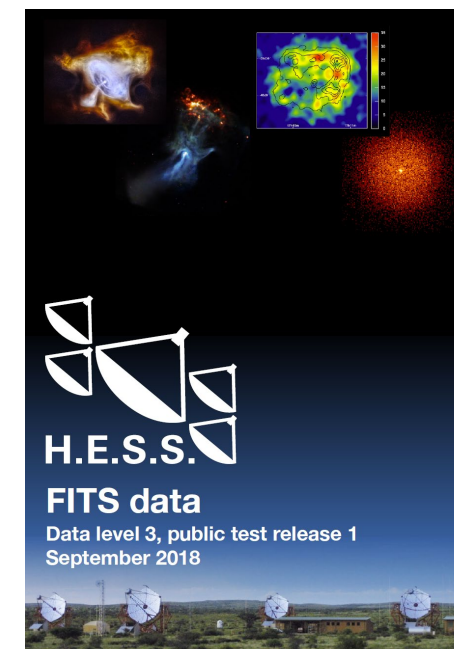
Structured FAIR Data Management in APP?

- **IMHO: it is still in its infancy!**
- **Though there are a few projects, e.g. in Helmholtz-programs, PUNCH, ErUM-Data, ACME**
- **Or dedicated service tools are going to be more FAIR, e.g. CORSIKA, Gammapy, AMPEL, KATRIN, etc.**



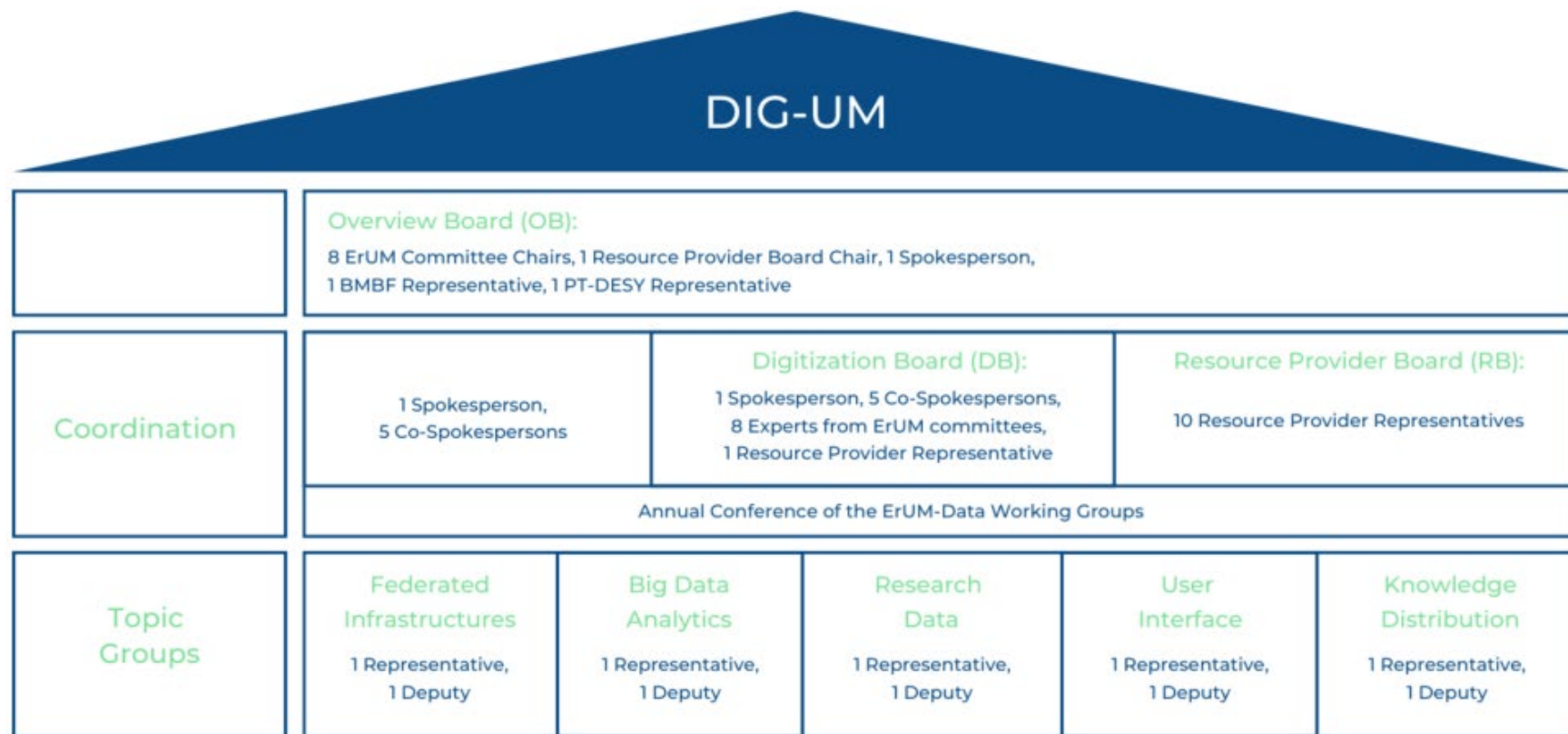
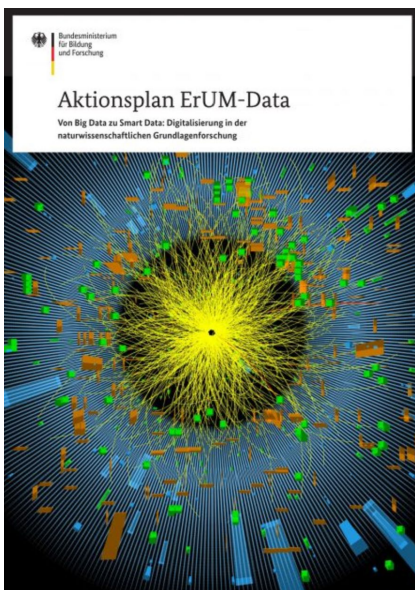
**One aspect: Open Data
(Open Data is not (only) Outreach)**

- **KCDC, Auger Open Data, H.E.S.S., IceCube,**




(2) Embedding / Services





DIG-UM unites the interests of the 8 research communities in the exploration of the universe and matter: Astroparticle Physics ([KAT](#)) Elementary Particles Physics ([KET](#)), Accelerator Physics ([KfB](#)), Research with Neutrons ([KFN](#)), Research with Synchrotron Radiation ([KFS](#)), Research with nuclear probes and ion beams ([KFSI](#)), Hadron and Nuclear Physics ([KHuK](#)) and the German Observatory Council ([RDS](#)).



The activities of the ErUM-Data-Hub are divided into 4 work areas:

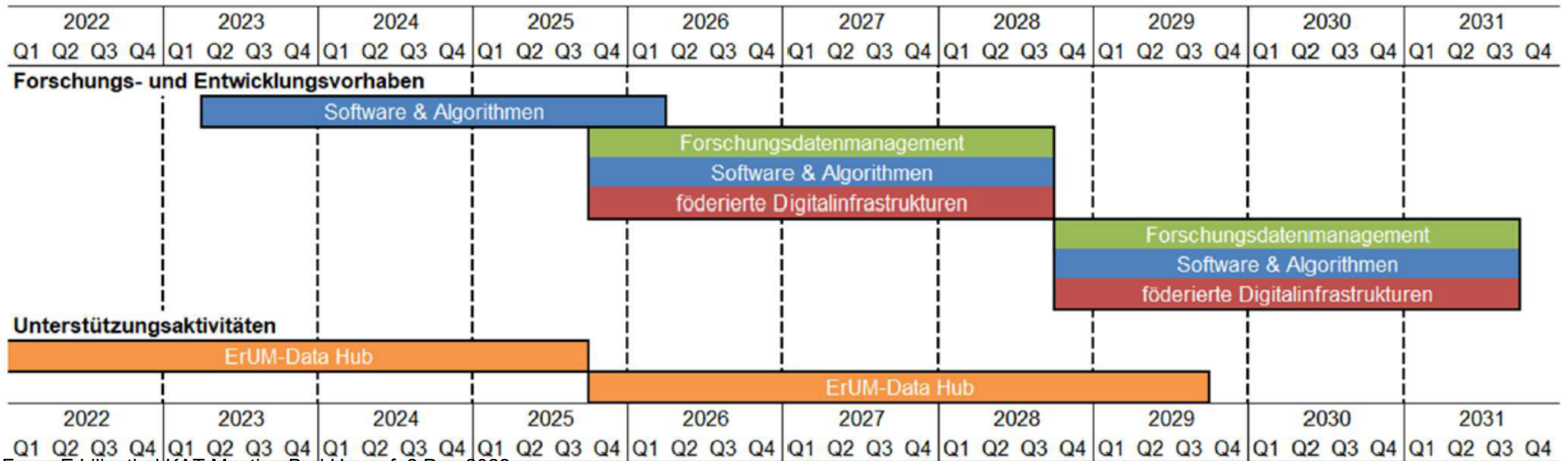
- Networking
- Exchange
- Digital competencies
- Communication

ErUM-Data call 2024

- Funding of the ErUM-Data Hub
- Calls and Funding of thematic R&D Projects
- „Software & Algorithms“ funding period 3/2023-2/2026
- „federated Infrastructures“ pilot project FIDIUM (KET/KHuK) 6/2022 3y

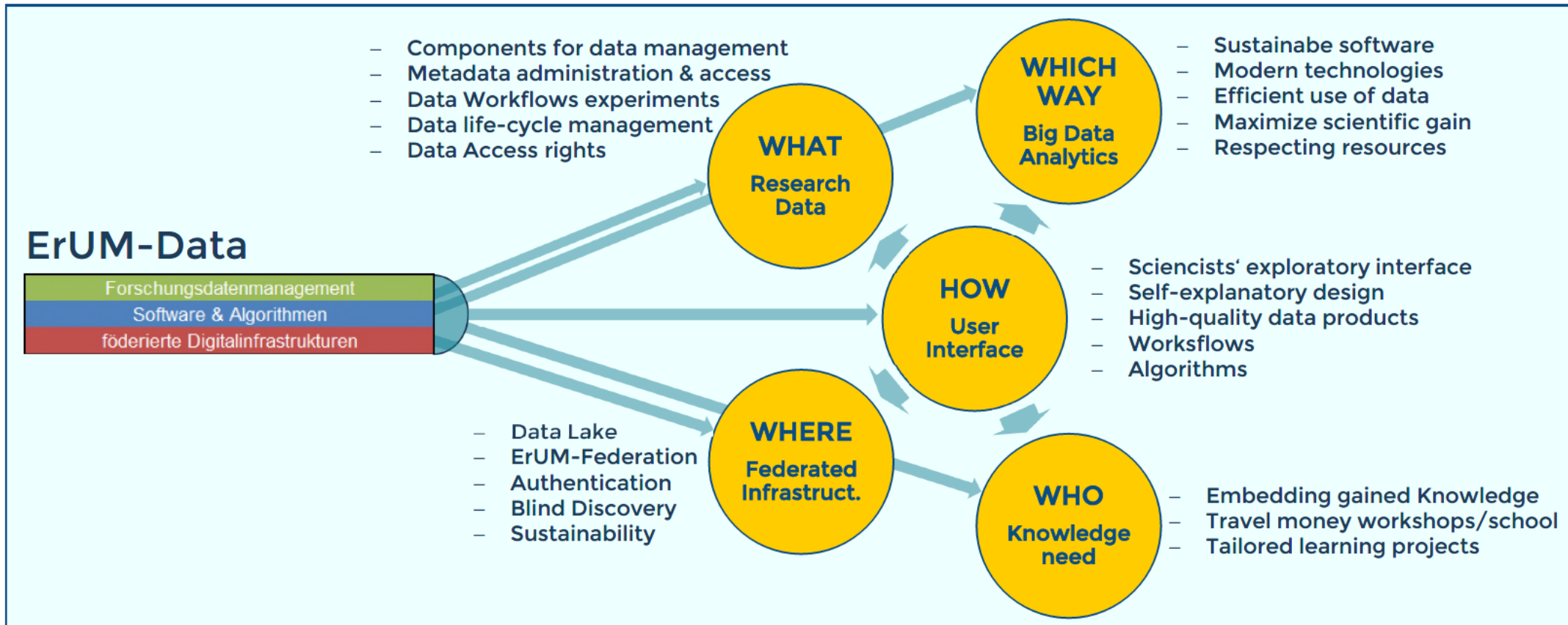
	No. joint projects (Verbünde)	No. of projects (Vorhaben)	(Requested) funding
Sketches (Skizzen)			74.5 Mio. €
Full proposals (Anträge)	33% of sketches		30.7 Mio. € 41% of sketches
funded	19% of sketches / 59% of full proposals	59% of full proposals	~17.5 Mio. € 23% of sketches 57% of full proposals

Software & Algorithms I



From: E.Lilienthal KAT Meeting Bad Honnef, 2 Dec 2022

Connection to ErUM-Data Themes



Martin Erdmann

KAT in the ErUM-Data Strategy Discussion

➤ ErUM-Data call

➤ Strategy Meeting with BMBF happened on 23+24 January 2024

➤ Call for projects 30. September 2024 → deadline 15. January 2025

https://pt.desy.de/bekanntmachungen/30092024_data/index_ger.html

➤ Start of the funding period will be 1 October 2025

➤ projects should be community overarching, and industry involvement is welcome

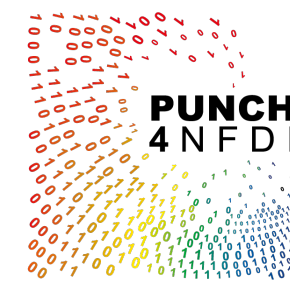
➔ Consortia are formed for all three areas (...could be more out of APP..)

➤ DIG-UM Annual Meeting 3-Dec-2024, 14:30

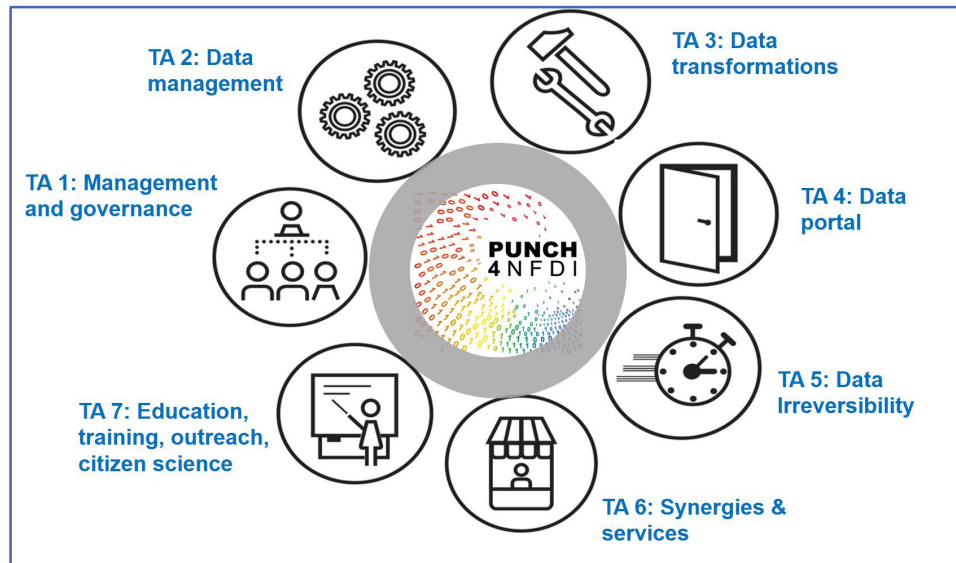
➤ ErUM-Data Hub organizes manifold workshops, see www.erumdatahub.de

➤ Connection ErUM-Data and EuCAIF (JENAS Network on AI) ; Next EuCAIF Conference scheduled for 15-20th of June 2025, Sardinia.

PUNCH4NFDI



- Particles, Universe, NuClei and Hadrons for the NFDI
- <https://www.punch4nfdi.de/>
- Proposal submitted 28.9.2020; Evaluation was on 10.12.2020
- 1 Oct 21: start of funding (reduced by 30% compared to proposal; 60 months / 16.6 Mio€)
- September 2024: evaluation report submitted
- February 2025: evaluation in Bonn
- August 2025: proposal submission for PUNCH 2.0

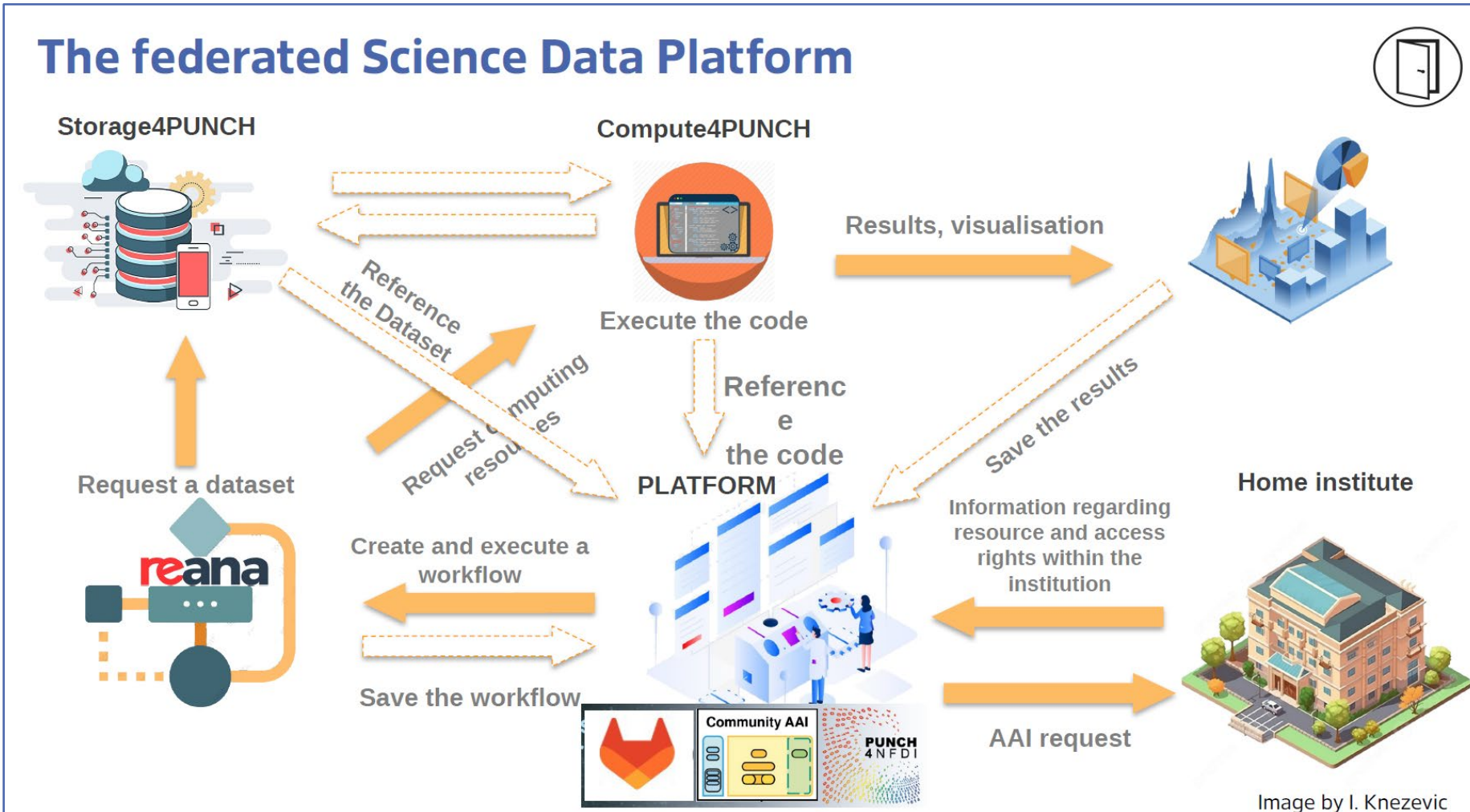


The PUNCH4NFDI Consortium

- Thomas Schörner (DESY, chair of consortium)
- Andreas Haungs (KIT, chair of statutes)
- Christiane Schneide (DESY, chair management)

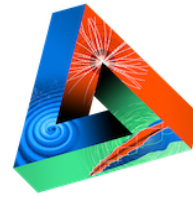
- Next general meeting 27-29 November in Bonn and online <https://indico.desy.de/event/45458>
- Thursday morning: **Public session: PUNCH4NFDI Community and User Meeting 2024**

The federated Science Data Platform



Discuss use cases for the SDP or parts of it (PUNCH toolbox for Data Management and Analysis) and attract new users!

JENA Computing Initiative



JENAA

Joint ECFA-NuPECC-APPEC Activities

- 1st JENA Symposium: 14-16 October 2019 in Orsay, France <https://jenas-2019.ijclab.in2p3.fr>
- 2nd JENA Symposium: 3-6 May 2022 in Madrid, Spain <https://indico.cern.ch/event/1040535>
- 3rd JENA Symposium: 8-11 April 2025 at RAL, UK <https://indico.cern.ch/event/1440480/>

- **One Topic: Future European Federated Computing**

Initiation started by workshop <https://agenda.infn.it/event/34738/>

Target: European white paper on (ENA) computing as input for the next JENA Symposium 2025 to discuss with representatives of funding agencies

Dedicated working groups (to look deeper) on five areas:

- HPC: https://indico.scc.kit.edu/e/JENA_computing_wp1/
- Software: https://indico.scc.kit.edu/e/JENA_computing_wp2/
- Data Management: https://indico.scc.kit.edu/e/JENA_computing_wp3/
- ML & AI: https://indico.scc.kit.edu/e/JENA_computing_wp4/
- Training: https://indico.scc.kit.edu/e/JENA_computing_wp5/



HORIZON-INFRA-2023-SERV-01-02 (domain: Astronomy & Astroparticle physics)

- Topic: **better access of users to RI services to advance frontier knowledge**
- Consortium: 41 partners, 15 countries, >30 research infrastructures
- Kick-off Meeting in Paris 16-17/9/2024
<https://indico.in2p3.fr/event/33636/>



(3) Computing and Data: in Roadmaps



How treated in neighboring roadmaps ?

- In RDS Denkschrift (draft):
 - Dedicated chapter on “High Performance Computing and e-science”,
 - in nearly all thematic topics hints that it is very important to have (federated / large-scale / green / high-performance) computing
- In ESPP 2020 (brochure):
 - Large-scale data-intensive software and computing infrastructures are an essential ingredient to particle physics research programmes
 - Further development of internal policies on open data and data preservation should be encouraged, and an adequate level of resources invested in their implementation. (FAIR not mentioned)
- In NuPECC Long Range Plans 2024 (draft):
 - Chapters for Open Science and Data (FAIR) / Quantum computing (QC) in nuclear physics / Machine learning (ML) and artificial intelligence (AI) in nuclear physics

Computing and Data Policies

APPEC requests all relevant experiments to continue to have their computing requirements scrutinised. APPEC will engage with the particle physics and astronomy communities to secure a balance between available European computing resources and needs for now and into the future. Appropriate training in data science should be provided for astroparticle physicists.



@ Connie Schneider / Unsplash

Open Science and Citizen Science

APPEC encourages the use of data format standards to facilitate data access between experiments. APPEC encourages funding agencies and publishers to support coherent Open Access publication policies. APPEC encourages making data publicly available as much as possible according to the FAIR principles. APPEC encourages citizen science to engage the public, while at the same time increasing the scientific capabilities of experiments.



Topic in current roadmaps:

- **KAT+RDS preparation for BMBF Strategy 2022:**

KAT und RDS empfehlen die Förderung von Entwicklungen, die Computing-Infrastrukturen und Softwarelösungen für die zukünftigen Projekte ertüchtigen und dabei gezielt Argumente der Nachhaltigkeit berücksichtigen.

Auch in der Astroteilchenphysik fallen Datenmengen im Petabyte-Bereich an; der Übergang zu einem federated computing und zu open and FAIR data ist im Gang, aber noch nicht abgeschlossen.

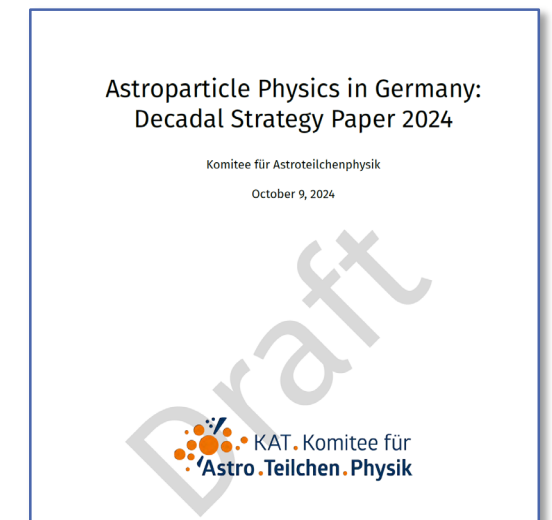
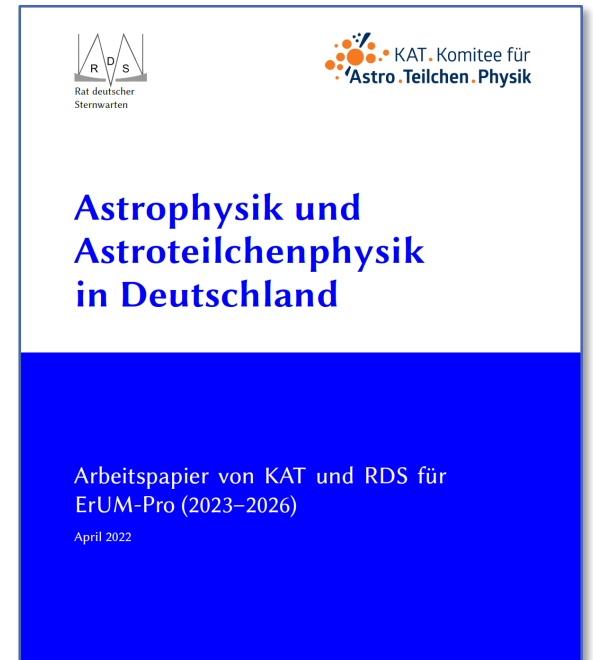
- **In current version of the KAT roadmap:**

4.6 Computing

The data

5.5 Research Data as a Cultural Heritage

The data produced by astroparticle physics experiments is not only a scientific asset but also a cultural heritage for future generations. The preservation and open accessibility of this data ensure that discoveries made today will continue to inform scientific exploration and contribute to a shared understanding of the universe.



ErUM-Prisma Trialogue (Sustainability)

- Working group „Sustainability in Computing“
- Recommendations in 5 categories:
 - The **cultural change** to sustainable computing in all areas of ErUM research should be actively supported.
 - The energy and resource efficiency of **IT infrastructures** should be presented transparently and must continue to be increased.
 - In the **life cycle analysis** of the required digital devices, the total costs must be optimized
 - Huge data streams management requires development of corresponding **algorithms and processes** for a FAIR and therefore sustainable data analysis and data reduction
 - **Software development**, which plays a central role in the sustainable use of resources, is to be promoted more intensively



(4) Discussion



Questions / Discussion

'FAIR' Digitalization

- How to treat topic “Computing and Data” in our KAT Roadmap
- What are the inherent needs for APP, in Germany: KAT
- Do we need a severe update on computing requests
- How concrete shall be the connection Digitalization & Sustainability
- Do we have to mention Quantum Computing
- What is our role in PUNCH, ErUM-Data, etc.
- What is the role of DZA in digitalization of APP
- your suggestions / questions?

...a little off-topic: Do we face the trap of justification vs. funding?

Earmarked funding vs. intrinsic research?

Do we want „digitalization for the research field“ or „digitalisation research with possible application for the research field“ ?

Do we want „sustainable research“ or „sustainability research“ ?

