Big-Data-Analytics

Preparation for Strategy Meeting

DIG-UM

DIG-UM Annual Meeting 2023

Thomas Kuhr, Jan Steinheimer

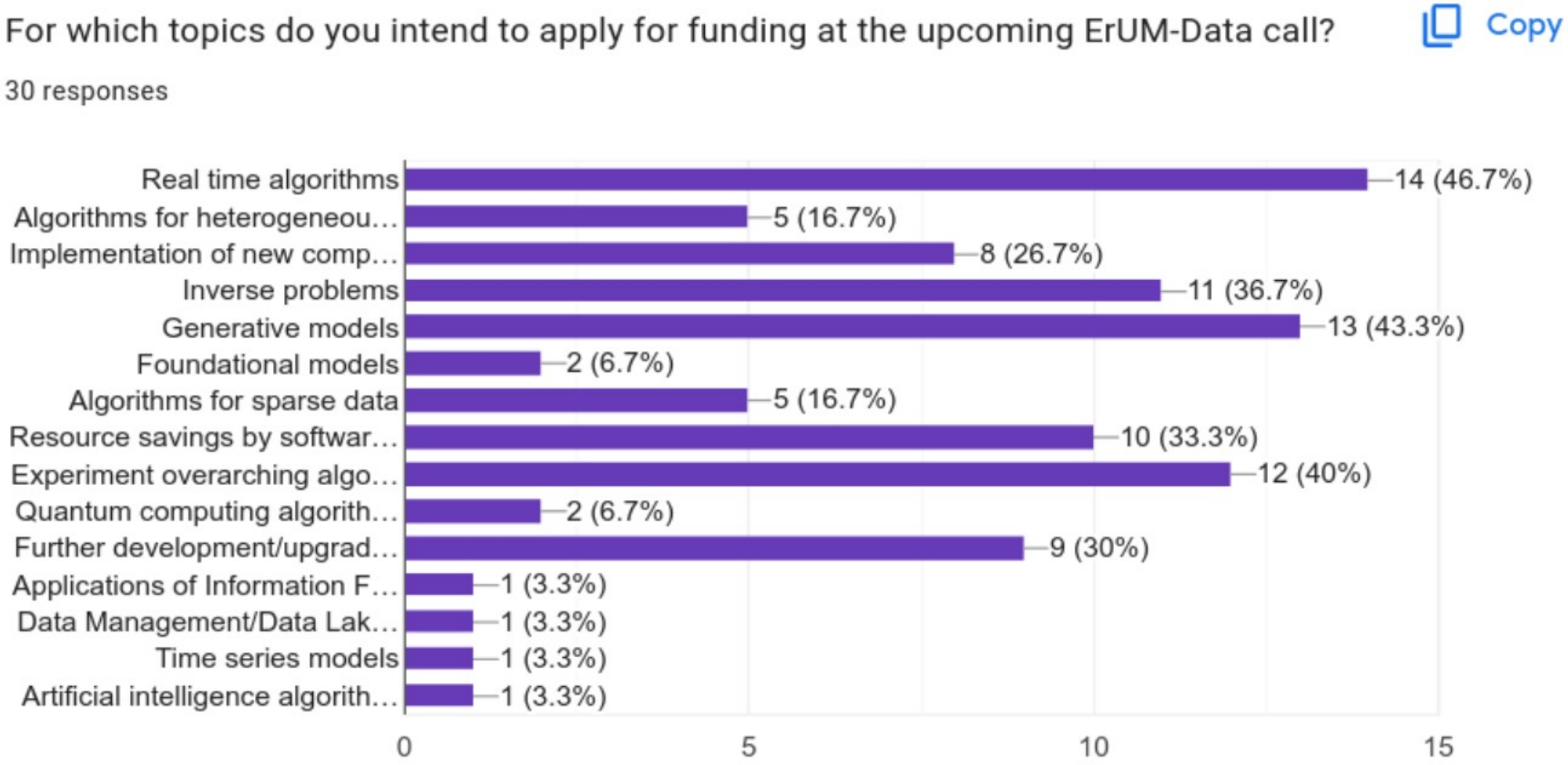
Proposed Outline of Talk at Strategy Meeting

- **Who are we? Organization of BDA topic group**
- **What are the challenges?**
- **Overview of running projects**
- **Examples of success stories (input from KISS received, more input welcome)**
- **Topics the communities want to address**
- **Further considerations**

Challenges

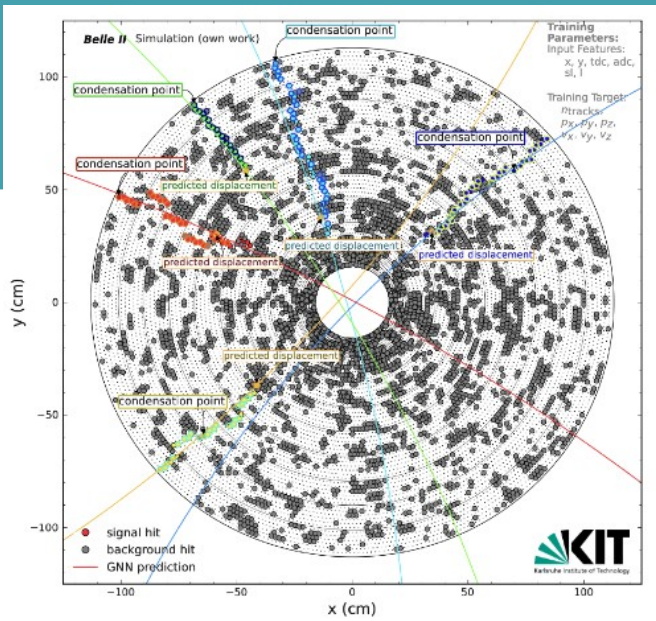
- Increasing data rate and volumes
- Sparse, specific data
- Required fast decisions
- Algorithms for heterogeneous computing
- Adoption of advanced technologies and methods
- Sustainability of research and environment

Survey Result on Topics

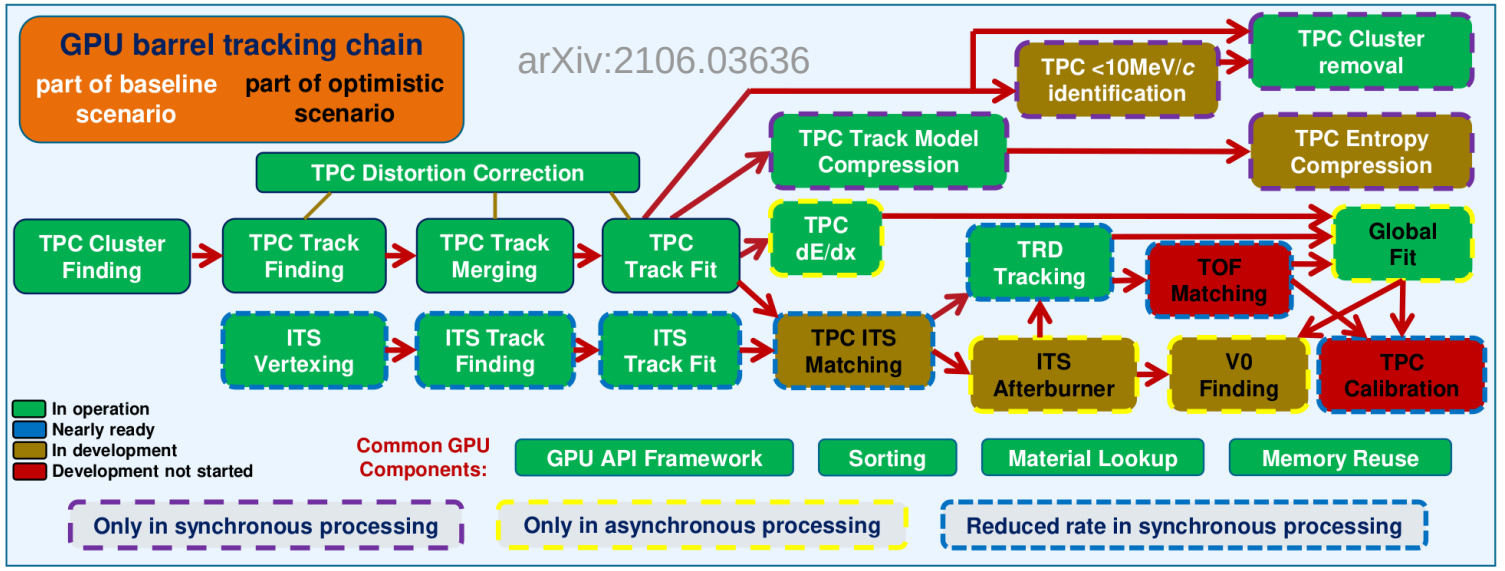


Real Time Algorithms

- **Fast decisions required for**
  - **Experiment optimization**
  - **Which data to store**
  - **Online analysis**
- ➔ **GPUs, FPGAs**
- ➢ **Algorithms for heterogeneous computing**
- ➢ **Implementation of new computational methods in dedicated hardware**

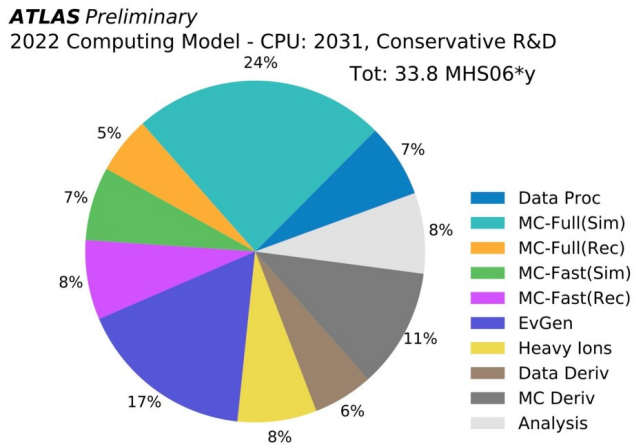


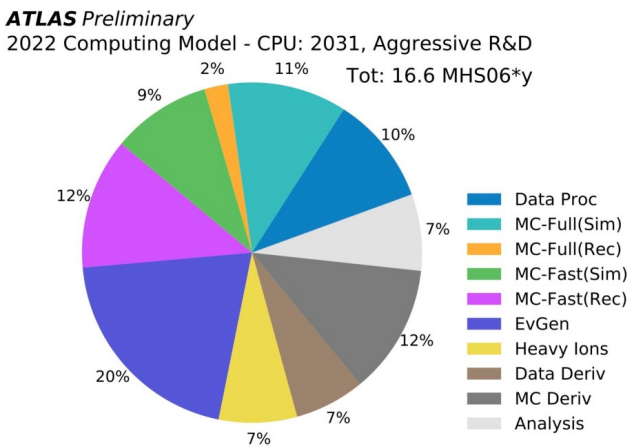
Example R&D: **GNN tracking for Belle II** with very low fake rates for displaced and low momentum tracks in high occupancy.



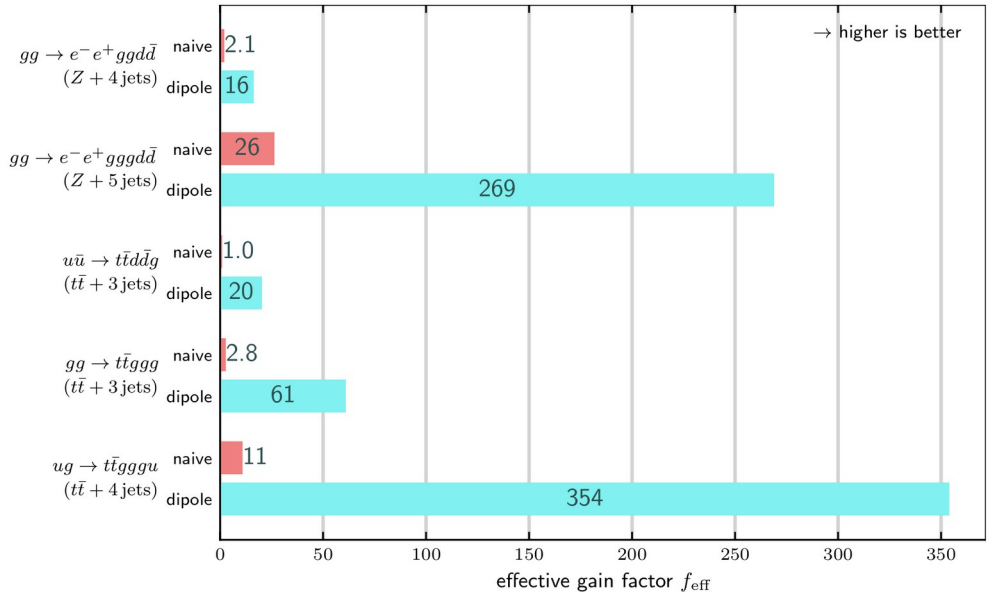
Generative Models

- **Simulations are often the only way to extract scientific results from measured data**
- **Huge resources are used for simulations**
- **Innovative methods to speed up simulations needed**





arXiv:2301.13562

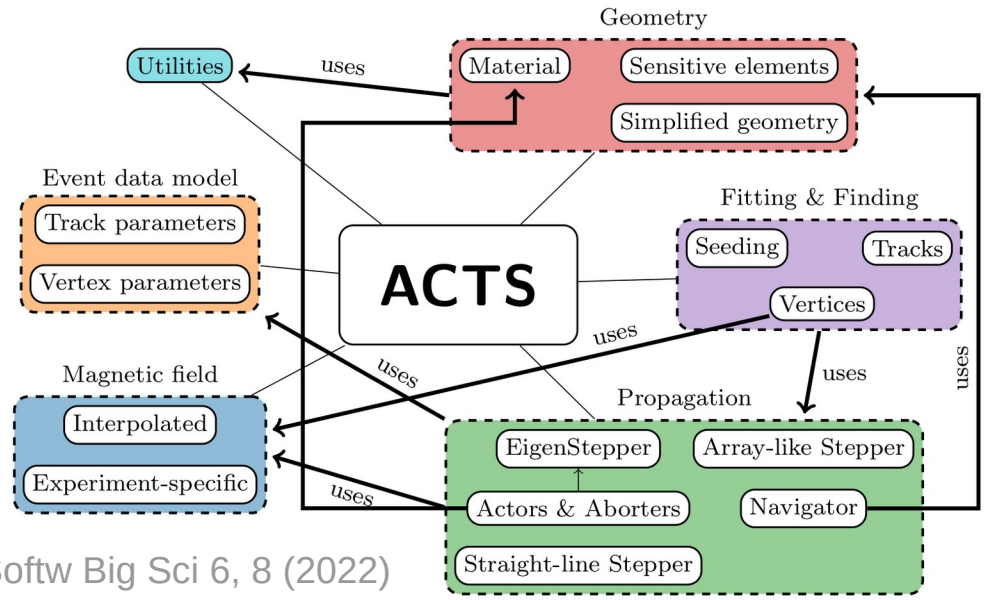


Inverse Problems

- **To be filled after workshop**

Experiment Overarching Algorithms

- **Some tools are already used successfully by multiple experiments, e.g. GEANT4 for detector simulation**
- **How far can we go in other areas?**
  - **E.g. ACTS for track recon-struction**



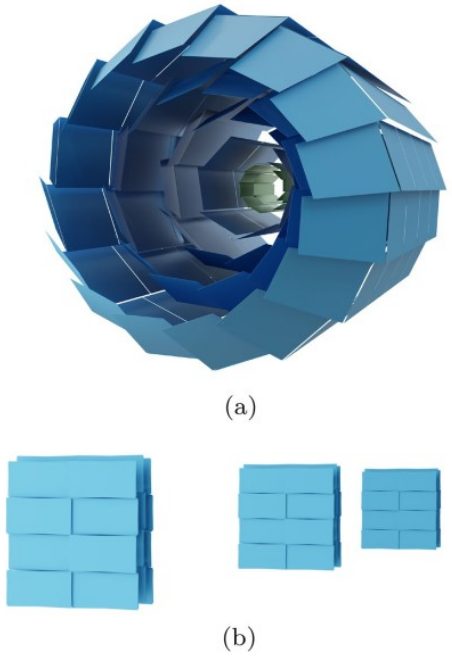


Fig. 10 Geometries of Belle II (**a**) and FASER (**b**) implemented in ACTS. Colors indicate the different subsystems

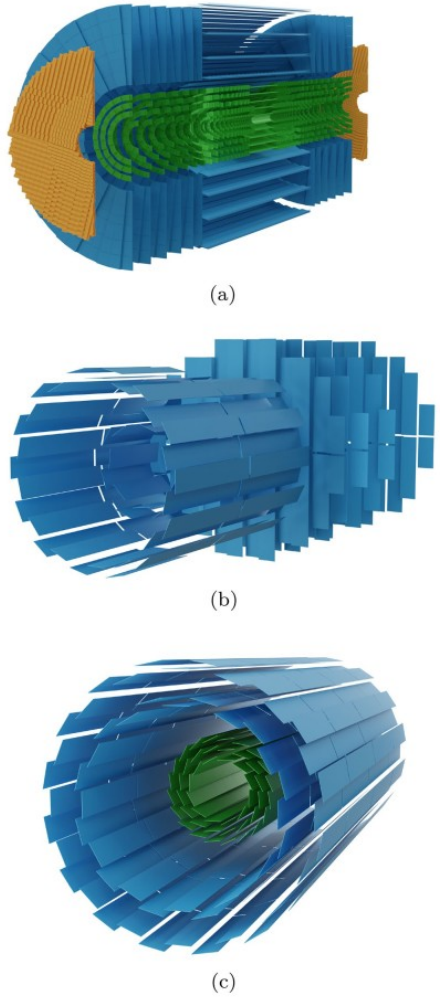


Fig. 9 The geometry of the ATLAS ITk (**a**), the PANDA silicon detector (**b**), and the sPHENIX silicon tracking detectors (**c**), implemented with ACTS. Colors indicate different subsystems; in the top image, the High Granularity Timing Detector (HGTD) [83] is shown in orange

Comput Softw Big Sci 6, 8 (2022)

More Topics

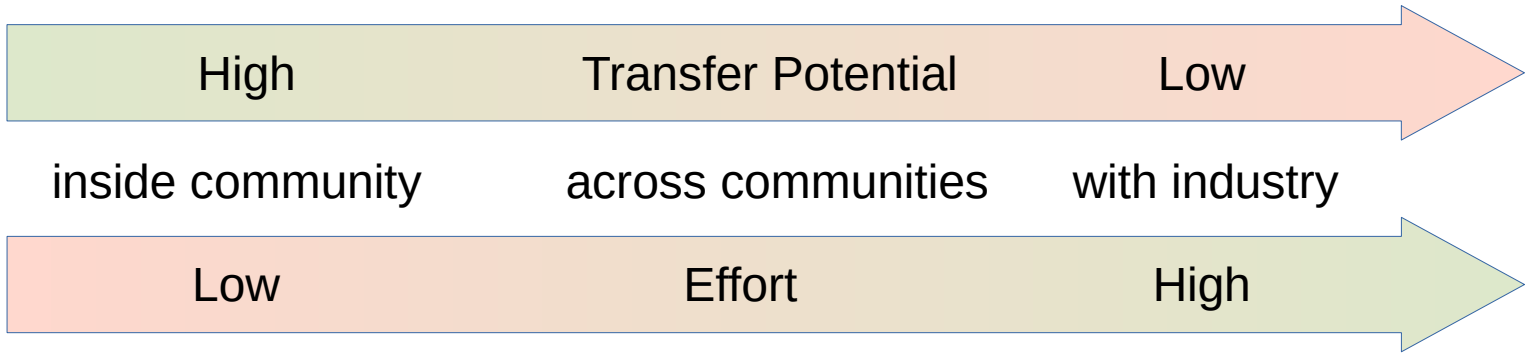
- **Further strong interest in:**
  - Resource savings by software optimization (→ heterogeneous hardware)
  - Further development/upgrade of existing software
  - Algorithms for sparse data

- **Other mentioned topics:**
  - Foundation models, quantum computing algorithms, AI assisted software development, applications of information field theory, time series models, AI algorithms for decay reconstruction

Further Considerations

- **Blue sky research vs. making developments usable long-term → both wanted**
- **Required contributions to experiments with experiment or community overarching potential in ErUM-Data**
- **Size and composition of collaborations according to needs, complementary collaborations, follow-up collaborations**
- **Transfer**

→ **Network**



Further Considerations (continued)

- Complementarity with NFDI
- International collaborations important
- Environmental sustainability (and sustainability of developed solutions) considered important
- Review with expertise in computing science and ErUM science wanted

Conclusions

- **Send us further input and material (nice figures with good explanation)**

Backup

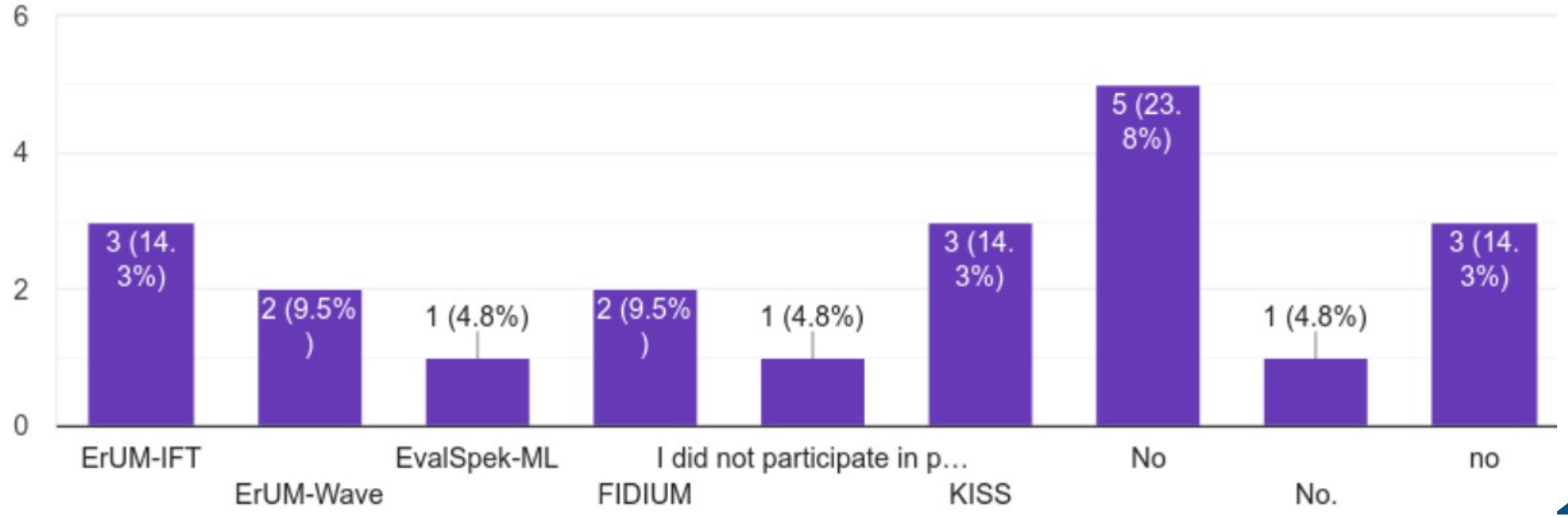
Survey Results

Would a project you apply for be based on results of a previous ErUM project (e.g. AISafety, ErUM-IFT, ErUM-Wave, EvalSpek-ML, KISS, KI4D4E, KI-Morph, OPAL-FEL, VIPR, 4D-KI-track, FIDIUM, other)? If yes, please give the name of the project(s).

Copy

21 responses

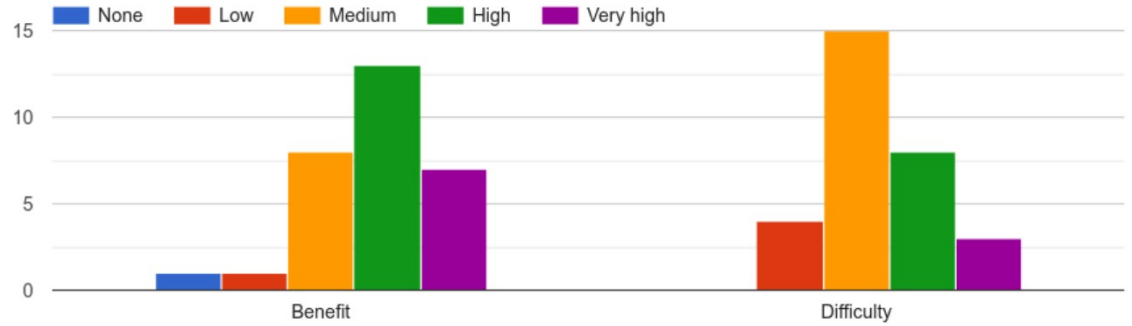
→ about one third said their project would be based on a previous project



Survey Results

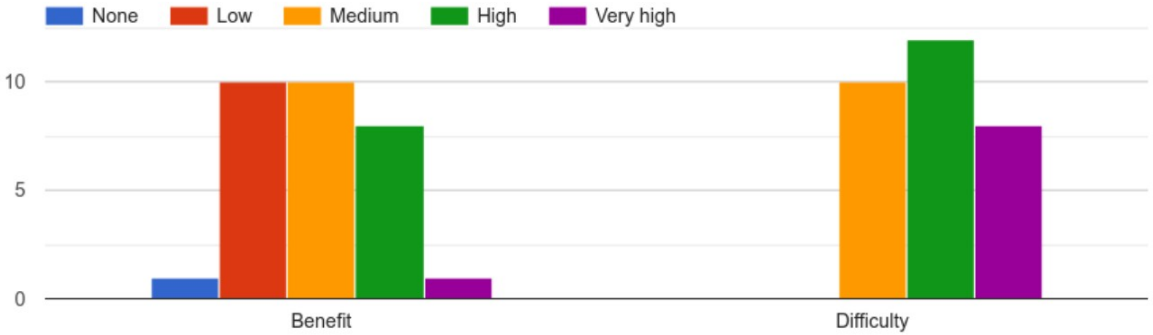
How do you see the collaboration across communities and the difficulty to find partners?

Copy



How do you see the collaboration with industry and the difficulty to find partners?

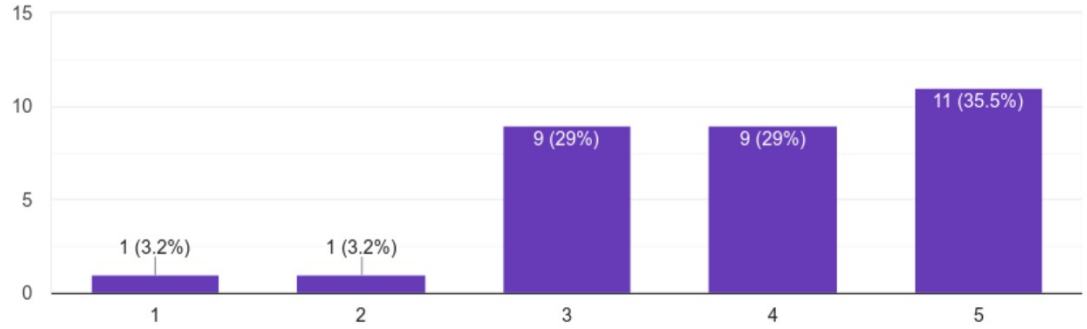
Copy



How important is the collaboration with international partners for the proposed projects?

Copy

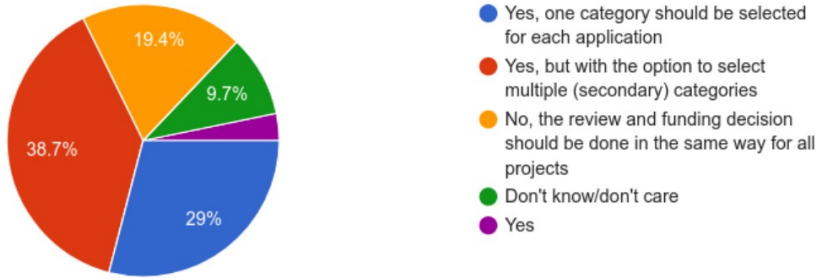
31 responses



Should applications and their review be separated by categories (federated infrastructures, software and algorithms, data management)?

Copy

31 responses



Survey Results

To the answer to "Should applications and their review be separated by categories" -> separating categories makes sense, but then it should be possible to submit separate applications to multiple targets. It would be a pity of an institution were forced to pick just one...

The communities accumulated a lot of know-how and solutions in their running projects. As a representative of a futuer project (Einstein-Telescope) I would love to see proposals of future projects with running projects to benefit from their extensice knowledge. We should not and do not want to reinvent the wheel with every new project!

Sustainability should play an important role

I support the change of the call away from the old format, in which applications along established community structures were preferred, to the new format that fosters interdisciplinary collaborations, and gives innovative approaches a chance.

1.) Honestly I see difficulties in the collaboration with industry due to the non-disclosure agreements. Results will not be open to public. 2.) If a funding decision for an application is negative, it would be very helpful (this time) to receive explanatory feedback and reasons for the decision. 3.) Differently from funding decisions in ErUM-Pro, funding decisions in ErUM-Data were only 100% positive or 100% negative. Partial funding was not considered. Question is, why there are those practically fundamental differences.

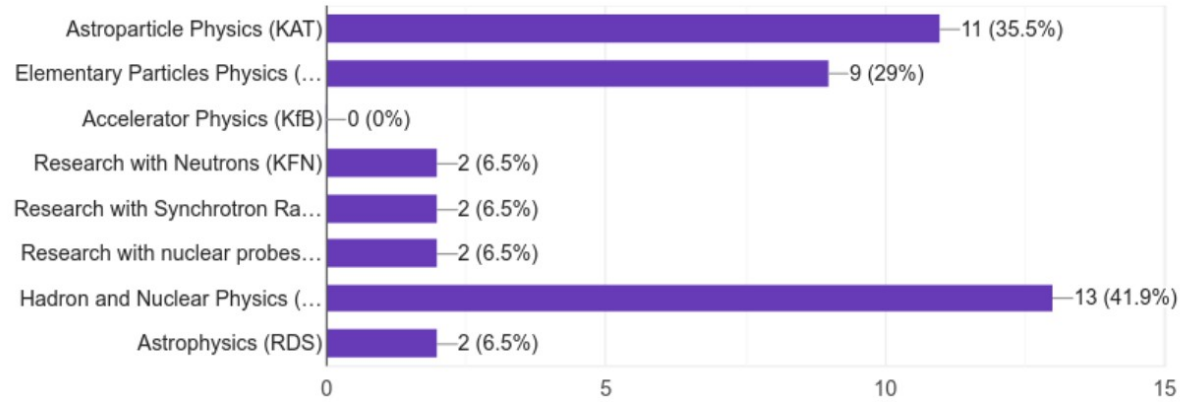
Separation of Digitization Needs for ErUM-Pro Experiments from General Developments in ErUM-Data needs to be clarified

Please include future projects in the calls. We want to learn from the running projects and don't want to reinvent the wheel.

What is your community?

Copy

31 responses



Who will take care of the long-term maintenance of the developed software?

Copy

30 responses