Big-Data-Analytics

Preparation for ErUM-Data Strategy Meeting
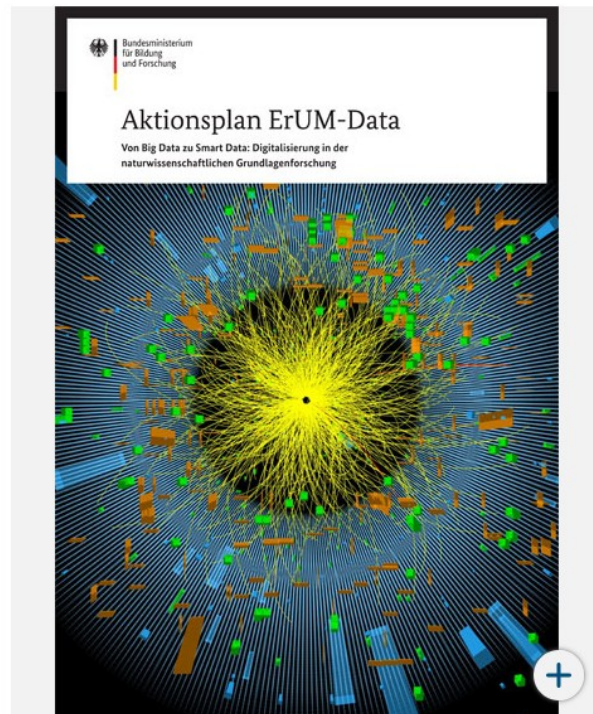
DIG-UM

ErUM-Data

ErUM-Data Projects So Far

- **Pilot project IDT-UM in ErUM-Pro funding period 2018-2021**
  - **Topics: computing, ML, tracking**
  - **Partners from KET, KHuK, KAT**

- **ErUM-Data funding period 2023-2026, with focus on AI and ML**
  - **10 funded projects: AISafety, ErUM-IFT, ErUM-Wave, EvalSpek-ML, KISS, KI4D4E, KI-Morph, OPAL-FEL, VIPR, 4D-KI-track**

DIG-UM

New ErUM-Data Call
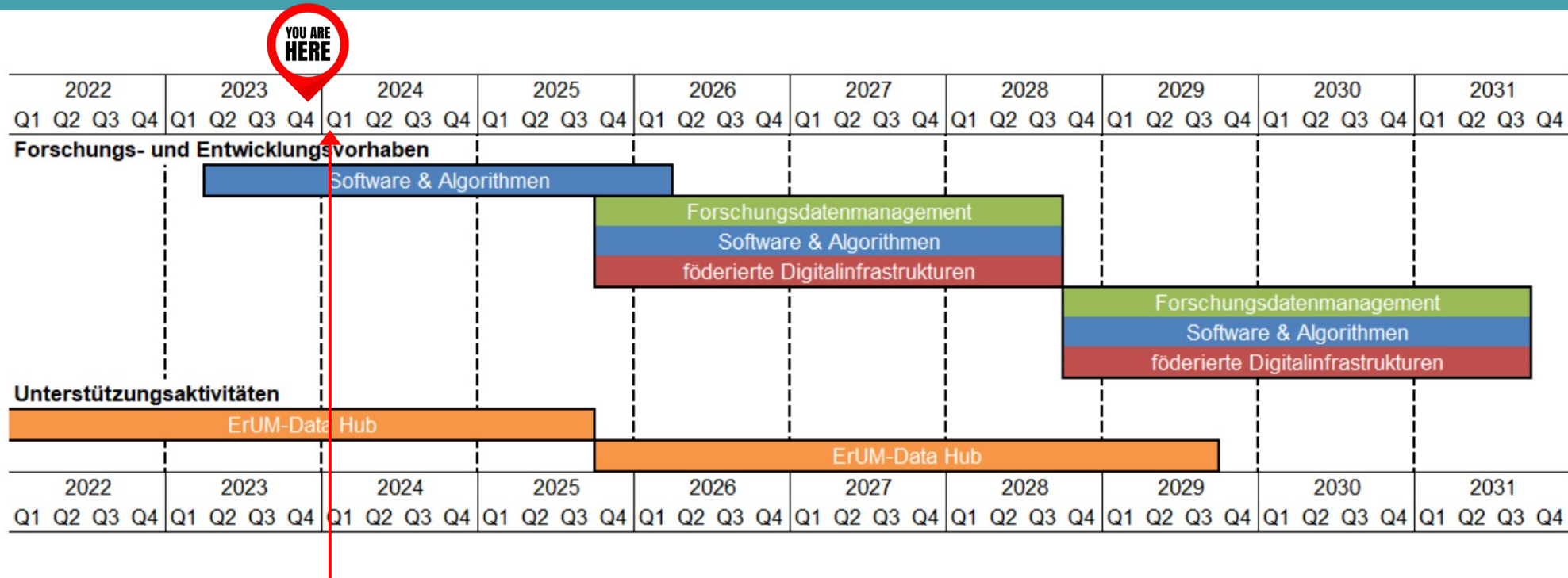
- **Strategy meeting on January 23/24 2024**
- **Talks by ErUM-Data-Hub and DIG-UM Topic Groups**

DIG-UM

DIG-UM

- **Organization of communities (KAT, KET, KfB, KFN, KFS, KFSI, KHuK, RdS)**

DIG-UM

	Overview Board (OB): 8 ErUM Committee Chairs, 1 Resource Provider Board Chair, 1 Spokesperson, 1 BMBF Representative, 1 PT-DESY Representative			
Coordination	1 Spokesperson, 5 Co-Spokespersons	**Digitization Board (DB):** 1 Spokesperson, 5 Co-Spokespersons, 8 Experts from ErUM committees, 1 Resource Provider Representative	**Resource Provider Board (RB):** 10 Resource Provider Representatives	
	Annual Conference of the ErUM-Data Working Groups			
Topic Groups	Federated Infrastructures 1 Representative, 1 Deputy	Big Data Analytics 1 Representative, 1 Deputy	Research Data 1 Representative, 1 Deputy	User Interface 1 Representative, 1 Deputy

(continued) Knowledge Distribution 1 Representative, 1 Deputy

➔ **https://erumdatahub.de/en/dig-um/**

DIG-UM

DIG-UM Topic Groups

- **Spokesperson: Martin Erdmann (KET, KAT)**

- **Federated Infrastructures: Markus Demleiter (RdS), Kilian Schwarz (KET)**

- **Big Data Analytics: Thomas Kuhr (KET), Jan Steinheimer (KHuK)**

- **Research Data Management: Astrid Schneidewind (KFN), Monica Valencia-Scheider (RdS)**

- **User Interface: Pierre Schnitzer (KfB), Tim Ruhe (KAT)**

- **Knowledge Distribution: Dirk Lützenkirschen-Hecht (KFS), Judith Reindl (KFSI)**

➔ **Mailing lists: https://wiki.erumdatahub.de/de/mailing-lists**

➔ **Annual meeting on Monday, December 11th**

DIG-UM

Proposed Outline of BDA Talk

- **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
(What to say on inverse problems?)**

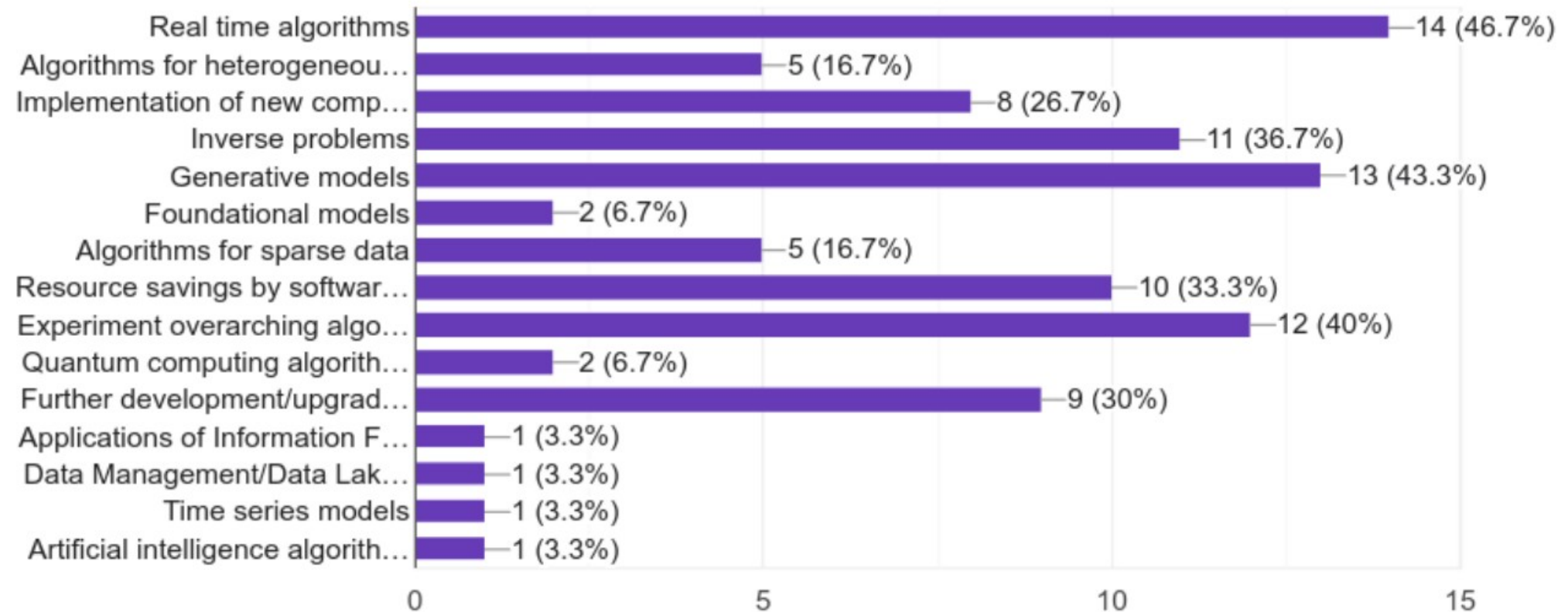- **Further considerations**

DIG-UM

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**

DIG-UM

Survey Results

For which topics do you intend to apply for funding at the upcoming ErUM-Data call?

30 responses

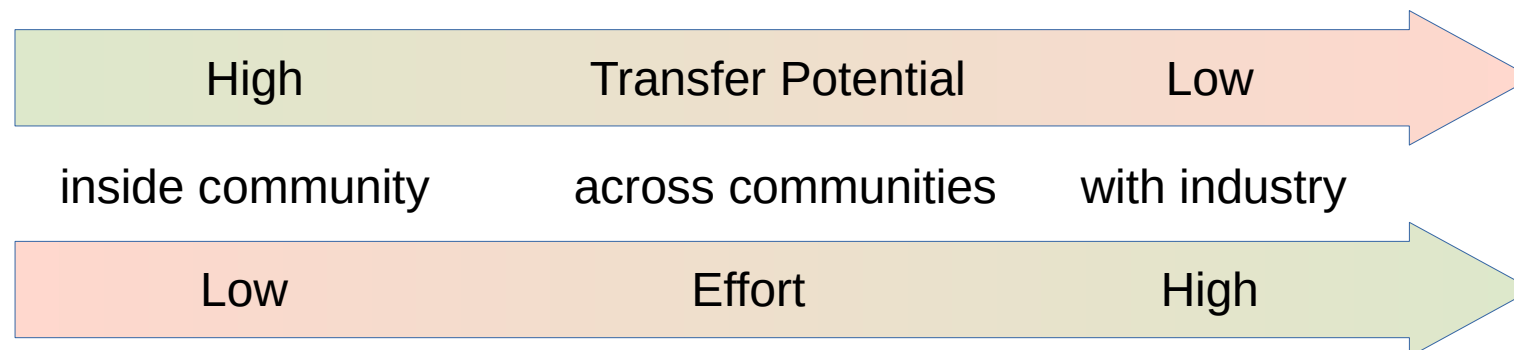| Topic	Responses
Real time algorithms	14 (46.7%)
Algorithms for heterogeneou...	5 (16.7%)
Implementation of new comp...	8 (26.7%)
Inverse problems	11 (36.7%)
Generative models	13 (43.3%)
Foundational models	2 (6.7%)
Algorithms for sparse data	5 (16.7%)
Resource savings by softwar...	10 (33.3%)
Experiment overarching algo...	12 (40%)
Quantum computing algorith...	2 (6.7%)
Further development/upgrad...	9 (30%)
Applications of Information F...	1 (3.3%)
Data Management/Data Lak...	1 (3.3%)
Time series models	1 (3.3%)
Artificial intelligence algorith...	1 (3.3%)

DIG-UM

Inverse Problems

- **To be filled after workshop**

DIG-UM

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? Complementary or follow-up collaborations?**

- **Transfer**

→ **Network**

High	Transfer Potential	Low
inside community	across communities	with industry
Low	Effort	High

DIG-UM

Further Considerations (continued)

- **Complementarity with NFDI**
- **International collaborations important**
- **Sustainability considered important**
- **Review with expertise in computing science and ErUM science wanted**

06.12.2023

DIG-UM

Conclusions

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

06.12.2023
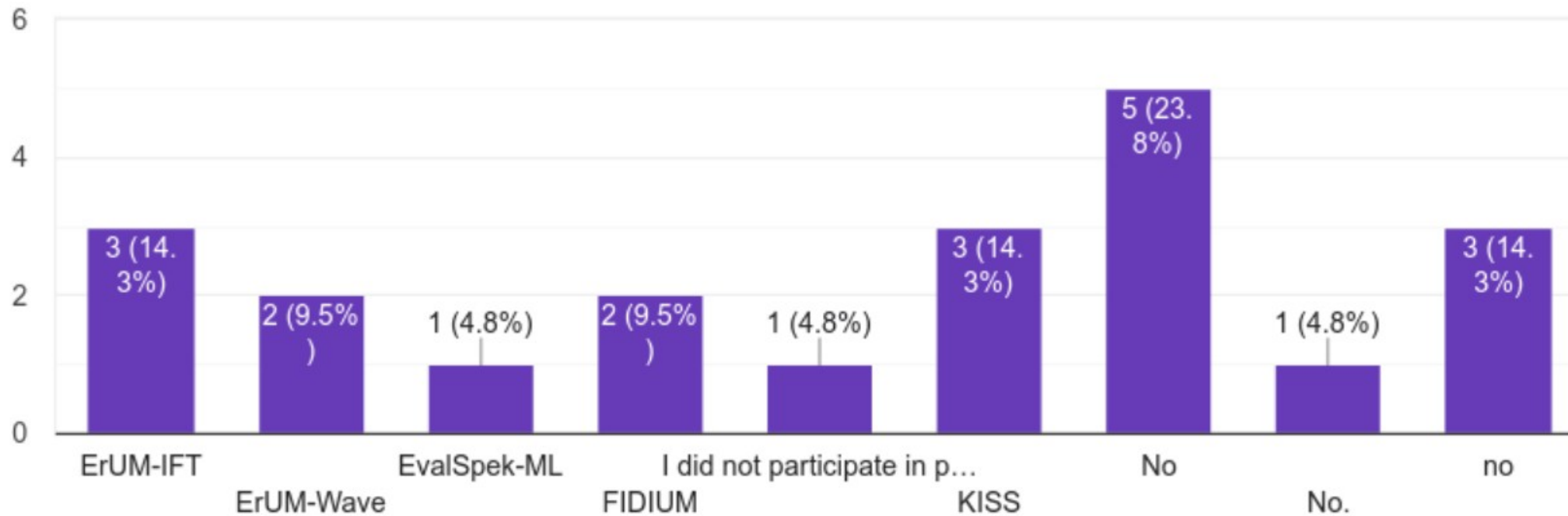
DIG-UM

Backup

06.12.2023

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).
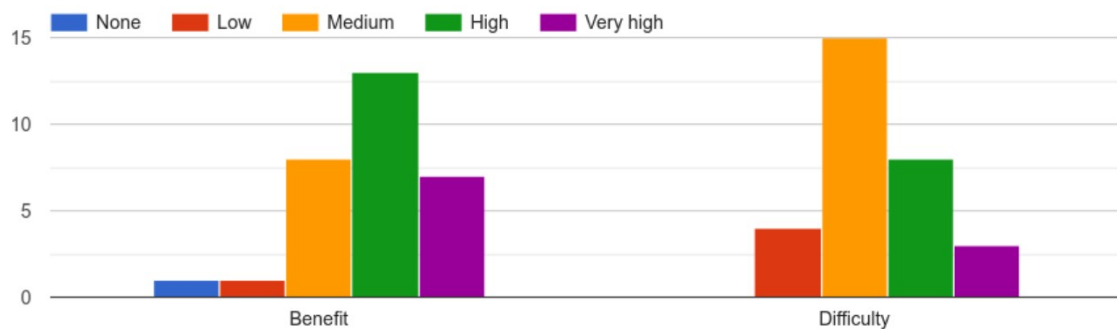
⧉ Copy

21 responses

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

06.12.2023
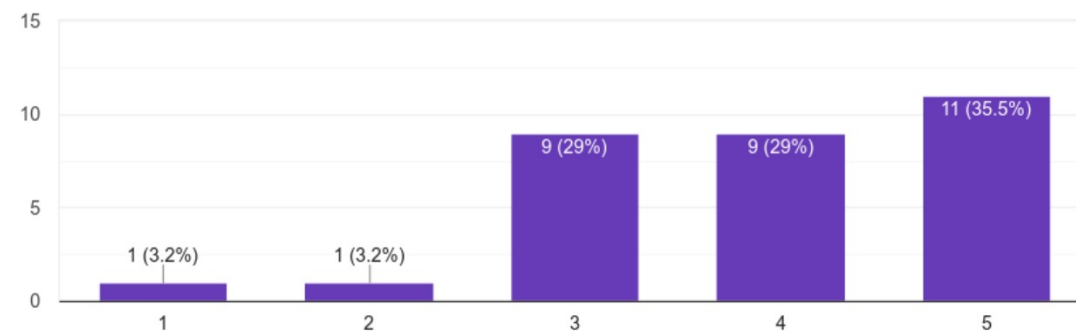
DIG-UM

Survey Results

How do you see the collaboration across communities and the difficulty to find partners? ⎘ Copy
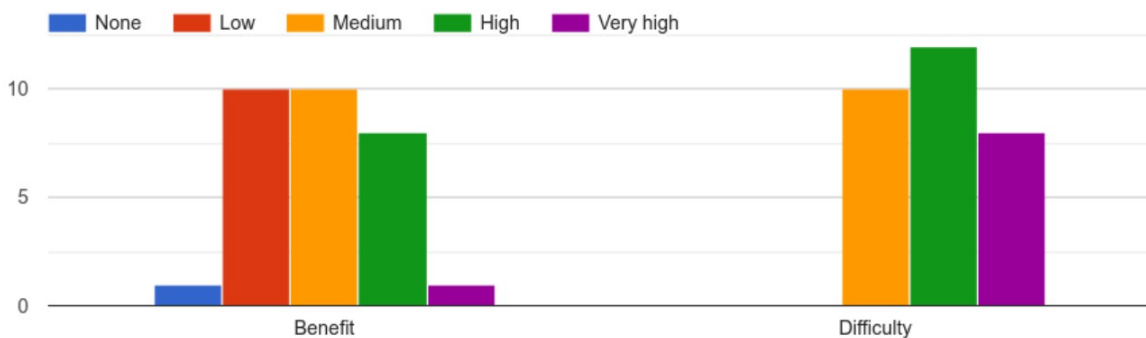
How important is the collaboration with international partners for the proposed projects? ⎘ Copy
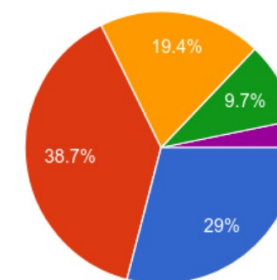
31 responses

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

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

31 responses

DIG-UM

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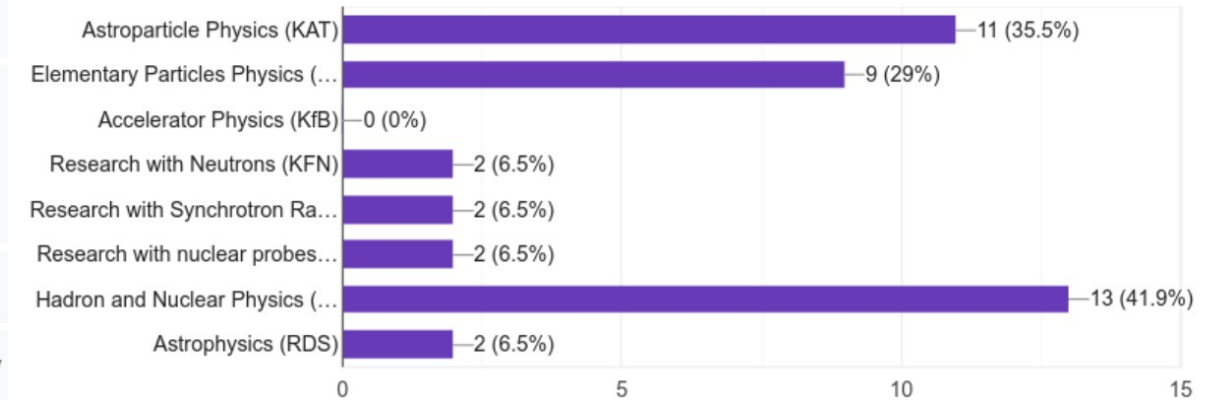
Separation of Digitization Needs for ErUM-Pro Experiments from General Developments in ErUM-Data needs to be clarified

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

Community	Responses
Astroparticle Physics (KAT)	11 (35.5%)
Elementary Particles Physics (...	9 (29%)
Accelerator Physics (KfB)	0 (0%)
Research with Neutrons (KFN)	2 (6.5%)
Research with Synchrotron Ra...	2 (6.5%)
Research with nuclear probes...	2 (6.5%)
Hadron and Nuclear Physics (...	13 (41.9%)
Astrophysics (RDS)	2 (6.5%)

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

30 responses

Option	Responses
Experiment / (international) coll...	19 (63.3%)
Research center	10 (33.3%)
Institute / university	10 (33.3%)
Industry	0 (0%)
Don't know	4 (13.3%)
A long-term maintenance is not...	2 (6.7%)
The user community of the soft...	1 (3.3%)
Max Planck Insitute for Astroph...	1 (3.3%)
This is a big problem for univer...	1 (3.3%)