

DIG-UM Annual Meeting 11-Dec-23,  
14:00-16:00, please register at

<https://indico.desy.de/event/42173>



# ErUM-Data & DIG-UM

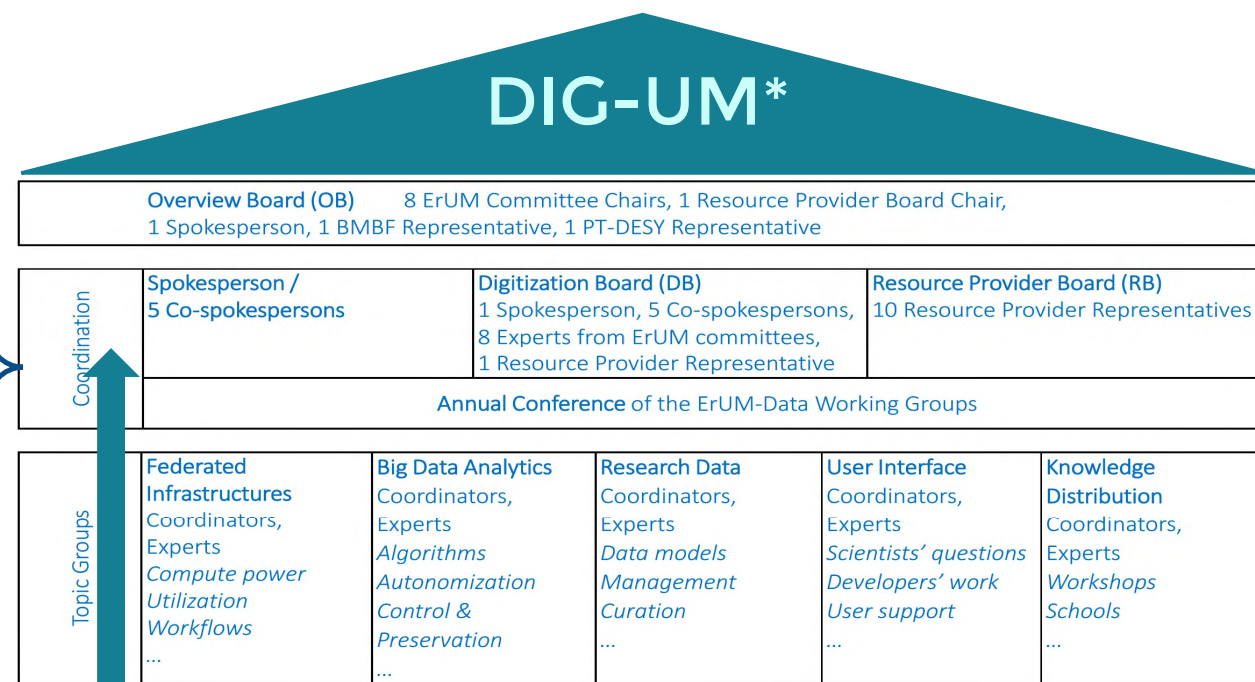
Martin Erdmann  
RWTH Aachen University  
7-Dec-2023

# Structure: Digital Transformation

## Community Self-Organization

### Digital Transformation

KAT  
KET  
KfB  
KFS  
KFSI  
KFN  
KHuK  
RDS



Spokesperson  
M.E.

## BMBF Funding

### ErUM-Pro

Project funding  
Astroparticle  
Astronomy  
LHC experiments  
Synchrotron  
sources  
Neutron sources  
Observatories  
...

### ErUM-Data

120 M€ / 10 years  
Project funding for  
Software & Algorithms  
...  
 ErUM-Data-Hub  
Digital Knowledge  
Agent ` 2,5M€ / 4 years

\*DIGital transformation in research on Universe & Matter

# ErUM-Data Software & Algorithms

## Hamburg/DESY Workshop 23.-24. Feb. 2023

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

|  |                    |               |
|--|--------------------|---------------|
| <b>Introduction to ErUM-Data from the PT-DESY</b><br>SR4, DESY | Marvin Berlinghof  | 14:10 - 14:30 |
| <b>Verbundvortrag - VIPR</b><br>SR4, DESY                      | Marina Ganeva      | 14:30 - 14:45 |
| <b>Verbundvortrag - KISS</b><br>SR4, DESY                      | Gregor Kasieczka   | 14:45 - 15:00 |
| <b>Verbundvortrag - KI4D4E</b><br>SR4, DESY                    | Sven Simon         | 15:00 - 15:15 |
| <b>Verbundvortrag - AISafety</b><br>SR4, DESY                  | Matthias Schott    | 15:15 - 15:30 |
| <b>Verbundvortrag - ErUM-IFT</b><br>SR4, DESY                  | Torsten Ensslin    | 15:30 - 15:45 |
| <b>Verbundvortrag - OPAL-FEL</b><br>SR4, DESY                  | Henrik Tünnermann  | 16:15 - 16:30 |
| <b>Verbundvortrag - EvalSpek-ML</b><br>SR4, DESY               | Sebastian Busch    | 16:30 - 16:45 |
| <b>Verbundvortrag - ErUM-WAVE</b><br>SR4, DESY                 | Prof. Conny Hammer | 16:45 - 17:00 |
| <b>Vorstellung Verbundantrag - aNNomalie</b><br>SR4, DESY      | Soeren Lange       | 17:00 - 17:15 |



# ErUM-Data-Hub @ Aachen

The Networking and Transfer Office  
serving Digital Transformation in Research  
on Universe & Matter in Germany

Martin Erdmann, Peter Fackeldey, Benjamin Fischer,  
Stefan Fröse, Judith Steinfeld, Angela Warkentin



SPONSORED BY THE



>440 Participants  
5 deep learning schools  
5 expert workshops  
4 train the trainer workshops  
100% recommendation

07.12.2023



Networking

Education



Deep Learning School

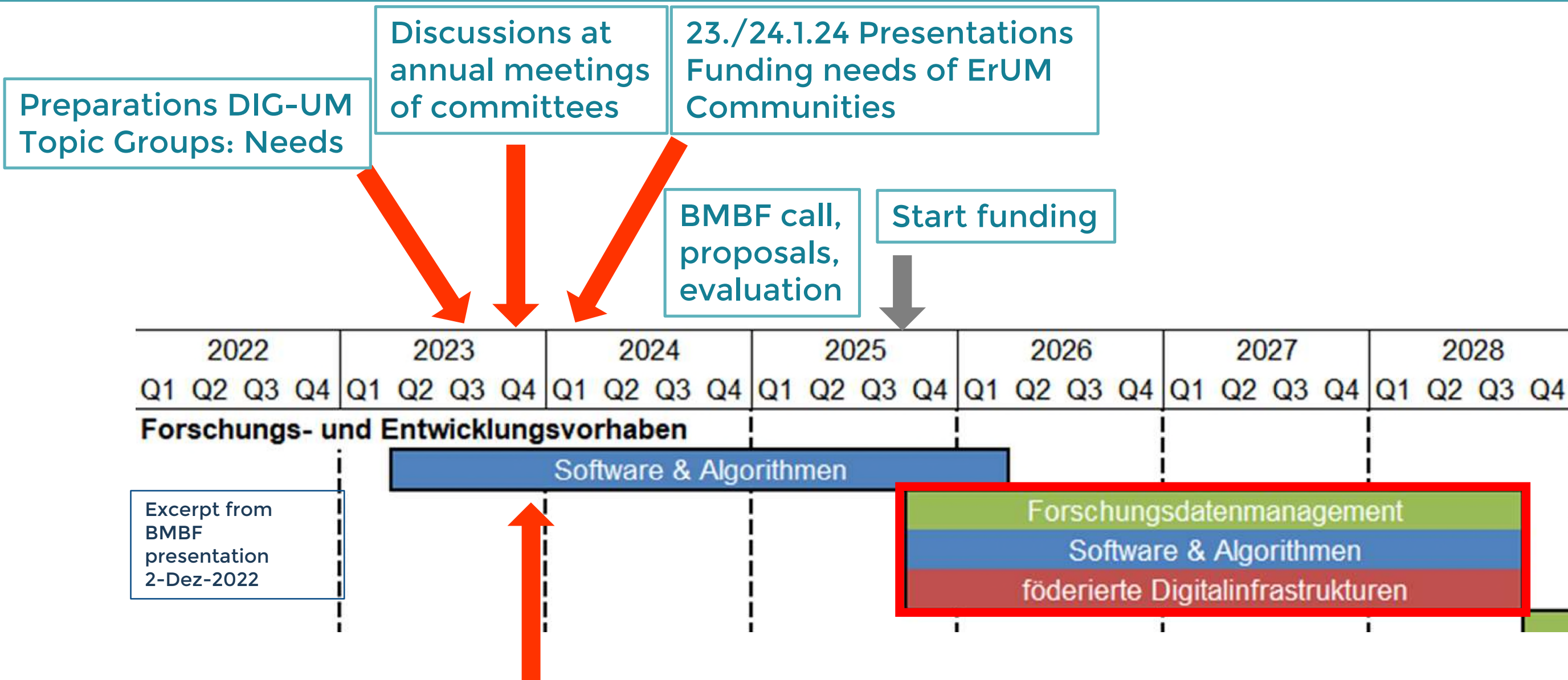


Transfer

Hannover  
Messe

# ErUM-Data Strategy Meeting 23/24-Jan-2024

## Discuss upcoming BMBF call for proposals



Annual meeting DIG-UM 11-Dec-2023, 14:00-16:00, followed by Overview Board meeting

07.12.2023

# ErUM-Data Strategy Meeting 23/24-Jan-2024

## Federated digital infrastructures

| DIG-UM  |   |   |   |  |
|---|---|---|---|--|
| Overview Board (OB)   |   | 8 ErUM Committee Chairs, 1 Resource Provider Board Chair  |   |  |
| 3 Spokesperson, 1 BMBF Representative, 1 FT DESY Representative |   |   |   |  |
| Committees  | Spokesperson / Co-Spokespersons   | Digitization Board (DB)   | Resource Provider Board (RB)  |  |
|   |   | 11 Spokesperson, 9 Co-Spokespersons, 8 Experts from ErUM committees, 1 Resource Provider Representative | 10 Resource Provider Representatives  |  |
| Work Groups   | Annual Conference of the ErUM Data Working Groups   |   |   |  |
|   | Federated Infrastructures<br>Coordinators<br>Experts<br>Computer power<br>Infrastructure<br>Workflows | Big Data Analytics<br>Coordinators<br>Experts<br>Algorithms<br>Automation<br>Control & Presentation     | Research Data<br>Coordinators<br>Experts<br>Data models<br>Management<br>Curation | User Interface<br>Coordinators<br>Experts<br>Scientists' questions<br>Developers' work<br>User support |

### All ErUM communities main challenge: Data avalanche by novel machines & experiments & observatories

#### Compute infrastructures

- Federated ErUM science cloud
- Energy efficient methodologies
- Container checkpoint & restore
- Transformation to NHR centers, data lakes

#### Data infrastructures

- Data federation (long-term)
- Combination of centers with substantial storage
- Connected by high bandwidth

#### Analysis facilities

- Dynamical setup
- Scalable workflows
- Heterogeneous resources
- Authentication & Authorization Infrastructure (AAI)

#### Simulation of novel computing models

- Data lakes, caches, sustainable computing

#### Monitoring & communication infrastructures

- Aim at sustainable computing energy provider - compute provider - user

#### Software Defined Networking

- Penalties for unscheduled transfers



# ErUM-Data Strategy Meeting 23/24-Jan-2024

## Big Data Analytics

| DIG-UM  |  |  |  |  |
|---|--|--|--|--|
| Overview Board (OB)   |  | 8 ErUM Committee Chairs, 1 Resource Provider Board Chair   |  |  |
| 3 Spokesperson, 1 BMBF Representative, 1 PT-DESY Representative |  |  |  |  |
| Committees  | Spokesperson / 8 Co-Spokespersons  | Digitization Board (DB)  | Resource Provider Board (RPB)  |  |
|   |  | 1 Spokesperson, 9 Co-Spokespersons, 8 Experts from ErUM committees, 1 Resource Provider Representative | 10 Resource Provider Representatives                                   |  |
| Work Groups   | Annual Conference of the ErUM-Data Working Groups                                    |  |  |  |
|   | Federated Infrastructures, Coordination, Experts, Computer power, Resource Workflows | Big Data Analytics, Coordination, Experts, Algorithms, Automation, Control & Presentation              | Research Data, Coordination, Experts, Data models, Management, Curator | User Interface, Coordination, Experts, Scientists' questions, Developers' work, User support |

### Algorithms

- Real time algorithms
- Inverse problems
- Generative modes (twins)
- Experiment overarching (tracking)
- Rapid classification for informed decisions, combined Accelerator – Experiment – Theory
- Outlier, Signal detection methods
- Sparse data

### Game-changing methods

- Transformers, large language models, foundation models
- Quantum computing

### Software

- AI-assisted development
- FAIR quality criteria
- Suited for heterogenous, parallel computing resources
- Sustainable in resource requirements

after web survey

# ErUM-Data Strategy Meeting 23/24-Jan-2024

## Research Data Management RDM

| DIG-UM  |   |  |  |   |
|---|---|--|--|---|
| Overview Board (OB)   |   | 8 ErUM Committee Chairs, 1 Resource Provider Board Chair   |  |   |
| 1 Spokesperson, 1 BMBF Representative, 1 FT DESY Representative |   |  |  |   |
| Committees  | Spokesperson /<br>8 Co-Spokespersons  | Digitization Board (DB)  | Resource Provider Board (RB)   |   |
|   |   | 11 Spokespersons, 9 Co-Spokespersons,<br>8 Experts from ErUM committees,<br>1 Resource Provider Representative | 10 Resource Provider Representatives   |   |
| Work Groups   | Annual Conference of the ErUM Data Working Groups   |  |  |   |
|   | Federated<br>Infrastructures<br>Coordinators,<br>Experts<br>Computer power<br>Infrastructure<br>Workflows | Big Data Analytics<br>Coordinators,<br>Experts<br>Algorithms<br>Automation<br>Control &<br>Presentation        | Research Data<br>Coordinators,<br>Experts<br>Data models<br>Management<br>Curation | User Interface<br>Coordinators,<br>Experts<br>Scientists' questions<br>Developers' work<br>User support |

ErUM connected to NFDI.

Requirements widely spread: high-end developments (KET, KHuK,...) - data management adapted within large scale facilities (KFS, KFN,...) - Virtual Observatory (RDS) -

### Systems of RDM

- Standard RDM solutions
- (User) Interfaces & programs
- Data catalogues
- Trainings & validation data for ML

### Systems of Metadata

- Management & documentation
- Transparency by general workflows from data reduction, analysis, (journal) publication

### Experiments & Steering

- Workflow systems for fully documented data treatment
- Monitoring, validation, visualization

### Data Lifecycle Management

- Replica management
- Strategies on availability of highly relevant data, deletion of uninteresting data
- Sustainable usage of compute resources

### Administration of access rights

- Systems to manage human & machine access to data, meta data
- Unification of authentication and access rights within ErUM



# ErUM-Data Strategy Meeting 23/24-Jan-2024

## User Interface

| DIG-UM  |  |  |   |  |
|---|--|--|---|--|
| Overview Board (OB)                               |  | 8 ErUM Committee Chairs, 1 Resource Provider Board Chair   |   |  |
| 3 Spokesperson, 1 BMBF Representative             |  |  |   |  |
| Committees  | Spokesperson / Co-Spokespersons                              | Digitization Board (DB)  | Resource Provider Board (RB)                            |  |
|   | 8 Co-Spokespersons   | 11 Spokespersons, 9 Co-Spokespersons, 8 Experts from ErUM committees, 11 Resource Provider Representatives | 10 Resource Provider Representatives                    |  |
| Annual Conference of the ErUM Data Working Groups |  |  |   |  |
| Work Groups                                       | Federated Infrastructures                                    | Big Data Analytics   | Research Data   | User Interface   |
|   | Coordinators, Experts, Computer power, Information Workflows | Coordinators, Experts, Algorithms, Automation, Control & Presentation                                      | Coordinators, Experts, Data models, Management, Curator | Coordinators, Experts, Scientists' questions, Developers' work, User support |
|   | Knowledge Distribution                                       | Coordinators, Experts, Workshops, Schools  |   |  |

All ErUM communities: Finding, using and connecting existing solutions must be simplified

Research Data Management

Was?

What is to be stored?

Big Data Analytics

Womit?

Which algorithms should be used to analyze the data?

User Interface

Wie?

How to access data and algorithms?

- Primary tasks:
- Enable easy access to data
  - Enable workflow management
  - Promote sharing of best practices
  - Interaction with experiment

- Workflow management
- Verifiable and reproducible workflows
  - Programming languages agnostic
  - Sufficiently documented
  - Low code/graphical programming
  - Comfortable Look and Feel

- Who-knows-what
- Person who knows who knows what
  - Knowledge to be decoupled from individuals & made discoverable

Federated Infrastructures

Wo?

Where should the data be stored?

Knowledge Distribution

Wem?

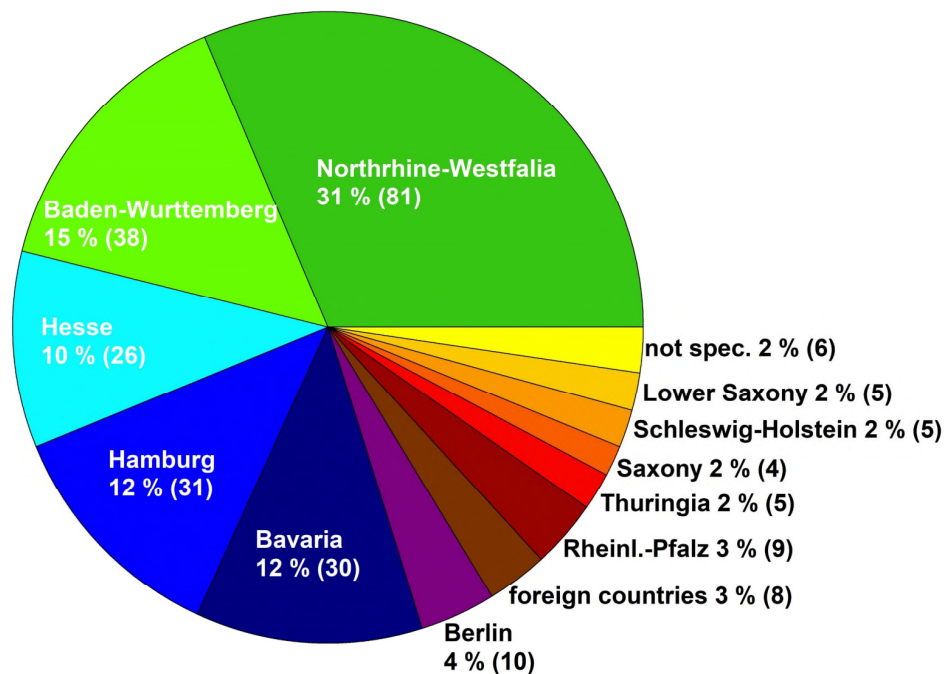
Who should we educate?

# ErUM-Data Strategy Meeting 23/24-Jan-2024

## Knowledge distribution

| DIG-UM  |  |  |   |   |
|---|--|--|---|---|
| Overview Board (OB)                               |  | ErUM Committee Chair, 1 Resource Provider Board Chair  |   |   |
| 1 Spokesperson, 1 BMBF Representative             |  | 1 Spokesperson, 1 BMBF Representative  |   |   |
| Committee   | Spokesperson / Co-Spokespersons                              | Digitization Board (DB)  | Resource Provider Board (RPB)                           |   |
|   | 1 Spokesperson, 9 Co-Spokespersons                           | 1 Spokesperson, 9 Co-Spokespersons, 8 Experts from ErUM committees, 1 Resource Provider Representative | 10 Resource Provider Representatives                    |   |
| Annual Conference of the ErUM Data Working Groups |  |  |   |   |
| Type/Group  | Federated Infrastructures                                    | Big Data Analytics   | Research Data   | User Interface  |
|   | Coordinators, Experts, Computer power, Information Workflows | Coordinators, Experts, Algorithms, Automation, Control & Presentation                                  | Coordinators, Experts, Data models, Management, Curator | Coordinators, Experts, Scientists' question, Developers' work, User support |
|   |  |  |   | Knowledge Dissemination, Coordinators, Experts, Workshops, Schools          |

### Common work with ErUM-Data-Hub team



Anchoring data science topics more firmly in universities

- Permanent anchoring in programs
- Teaching digital skills
- Start early Bachelor's program

Qualification of young scientists



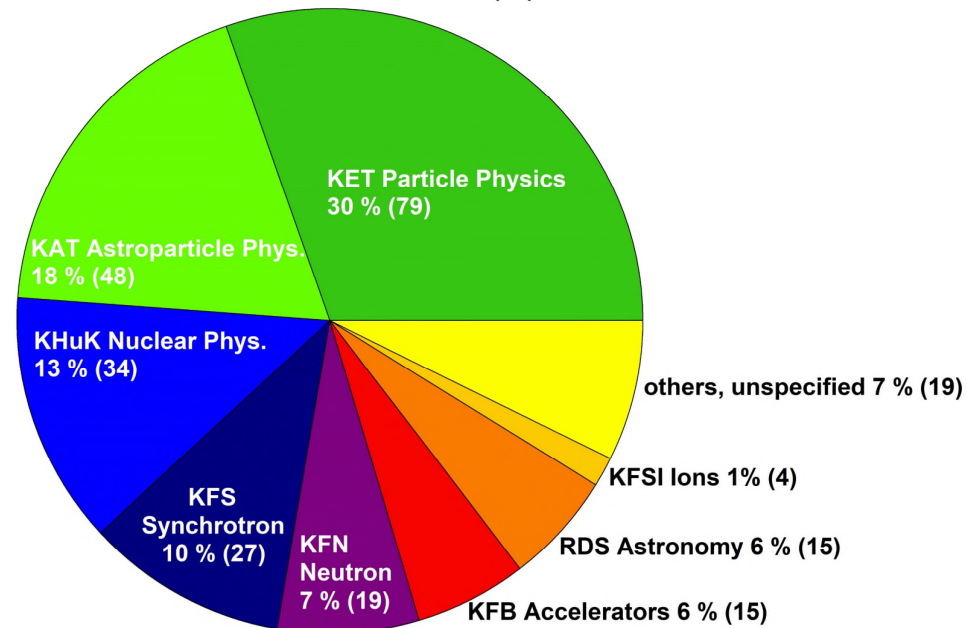
- Coaching offers
- Help for self-help
- Career mentoring

Innovative development of tailored-learning concepts

- Different concepts for knowledge dissemination are developed and made publicly available (e.g. problem-based learning)

Anchoring knowledge distribution: Outreach

- Funded project integrates concepts from community or realize its own concepts



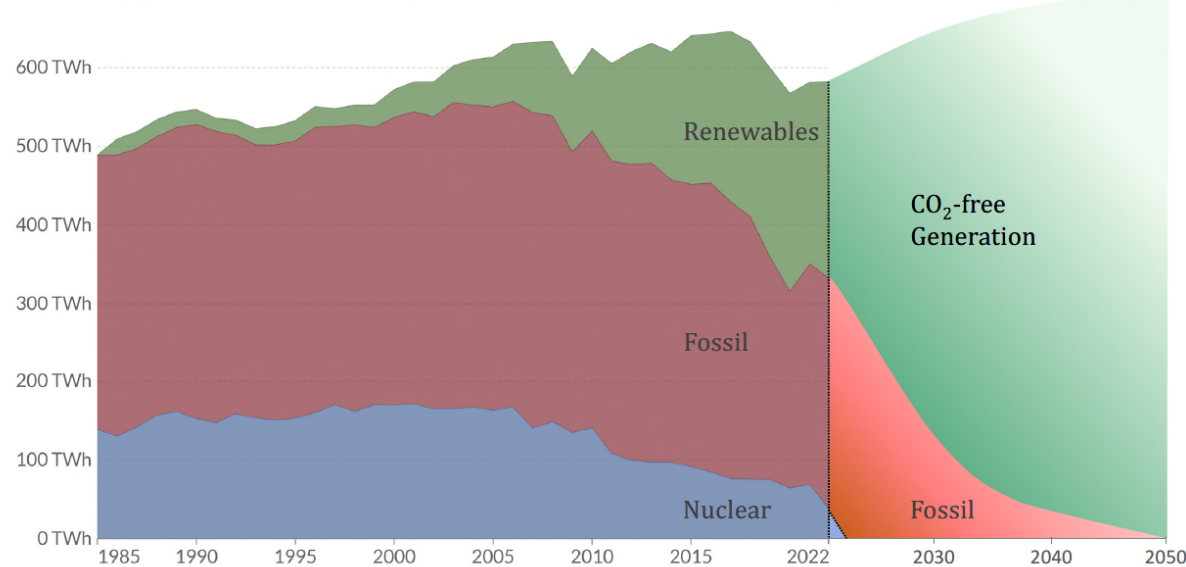
# Sustainability in ErUM Data

4-day Workshop May 23

Developing a culture for sustainable science is growing consensus in community

→ Awareness in daily research: balance *knowledge gain* with *resource usage*

Electricity production from fossil fuels, nuclear and renewables, Germany



Source: Ember's Yearly Electricity Data; Ember's European Electricity Review; Energy Institute Statistical Review of World Energy OurWorldInData.org/energy • CC BY

## Portfolio of measures

Short term

Medium term

Longer term

<https://arxiv.org/abs/2311.01169>

| Item   | Call-to-action   |
|--|--|
| Immediately or on <b>short time scale</b> with little effort these measures that can be implemented: |  |
| S1   | Raise awareness of the climate challenge at all levels.  |
| S2   | Disseminate knowledge of measures to address the challenge.  |
| S3   | Monitor and report energy consumption at job level.  |
| S4   | Consider carbon footprint for all investments and project plans.   |
| On a <b>medium time scale</b> of a few years the following measures can be realized:                 |  |
| M1   | Make data FAIR to promote reuse.   |
| M2   | Reduce and compress data having the anticipated scientific value of the retained information and the resource requirements in mind.  |
| M3   | Optimize the choice of storing against re-calculating intermediate results.  |
| M4   | Use workflow management to make processing FAIR.   |
| M5   | Make software FAIR and reliable by following good software development practices and ensuring sustainable support.   |
| M6   | Design software for optimized energy consumption and provide tools to measure it.  |
| M7   | Continue research on potential of AI or other new technologies for efficient use of resources, but balance gain of research action against resource consumption of these developments.                             |
| M8   | Monitor and report energy consumption at site and project level, provide information of the individual use per scientist/project/publication.  |
| M9   | Extend monitoring of resources beyond CO <sub>2</sub> e (water, material etc.).  |
| M10  | Train scientists in good practices.  |
| M11  | Regularly review and update the CO <sub>2</sub> e reduction plan.  |
| M12  | Strive to become a role model at all levels and help to establish sustainability in everyday life.   |
| A <b>longer term</b> coordinated planning is required for the following measures:                    |  |
| L1   | Adjust computing in space and time to the availability of renewable energy, e.g. computing centers close to off-shore wind parks with a job scheduling using only or mainly the surplus available at a given time. |
| L2   | Develop software and middleware that can respond dynamically to the availability of energy.  |
| L3   | Optimize power usage effectiveness.  |
| L4   | Re-use of produced heat.   |
| L5   | Adjust hardware lifetime considering emissions due to procurement and operation.   |
| L6   | Include the resources needed for continuous IT support into project planning.  |

24.5.23 BMBF PRISMA Trialog with 6 working groups on „Nachhaltigkeit in der Forschung an Großgeräten:

Ressourceneffizienz & Zukunftssicherung“

# ErUM-Data Strategy Meeting 23/24-Jan-2024

## Open Questions

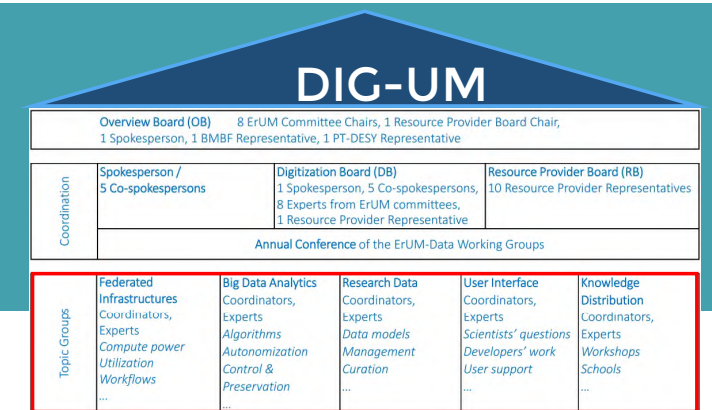
1. Interplay ErUM-Pro with ErUM-Data:
  - Experiment-related digitalization developments essential for their success
  - Cross-community creation & development of broadly applicable solutions
2. Matter, Universe, Particle:
  - Applications from two research fields mandatory?
3. Non-university partners mandatory for applications:
  - Industry partner?
  - Partner from large scale research infrastructure?
4. Federated Digital Infrastructures, Software & Algorithms, Research Data Management
  - Separate application fields or possible integration of all three areas in 1 application?
5. Recommendation for consortium size of application?



# Conclusions

- DIG-UM Community is well alive and very active
- Preparations for BMBF ErUM-Data Strategy meeting 23/24.1.24 well advanced
- Join DIG-UM Annual meeting 11.12.23 to observe current status
- Input your ideas to help bringing presentations in final shape
- Your search for consortium
  - Community Exchange by Erik Bründermann <https://indico.desy.de/event/28766>.
  - Topic groups plan for topical & local come-togethers (join mailing lists).
  - Talk to ErUM-Data-Hub Team how to find groups.

# DIG-UM Topic Group Mailing Lists



Register yourself at

<https://lists.rwth-aachen.de/postorius/lists/erum-data-federated-infrastructure.lists.rwth-aachen.de>

<https://lists.rwth-aachen.de/postorius/lists/erum-data-big-data-analytics.lists.rwth-aachen.de>

<https://lists.rwth-aachen.de/postorius/lists/erum-data-research-data-management.lists.rwth-aachen.de>

<https://lists.rwth-aachen.de/postorius/lists/erum-data-user-interface.lists.rwth-aachen.de> <https://lists.rwth-aachen.de/postorius/lists/erum-data-knowledge-distribution.lists.rwth-aachen.de>

**LISTNAME=erum-data-XXX**

**XXX=federated-infrastructure, big-data-analytics, research-data-management, user-interface, knowledge-distribution**

- To subscribe, send e-mail to **LISTNAME-request@lists.rwth-aachen.de** with single word “**subscribe**” in the **subject**. Leave the body of the e-mail empty.
- Subscription is open to all ErUM scientists (list owners can administrate).
- Sending e-mails to list: Anyone on the mailing list can send e-mails to the entire list. Use exactly the registered e-mail address. Other e-mail addresses can sent to the entire list after approval by the group convenor.
- To unsubscribe, send e-mail to **LISTNAME-request@lists.rwth-aachen.de** with single word “**unsubscribe**” in the subject. Leave the body of the e-mail empty.