Advancing Digital Transformation in Research on Universe and Matter in Germany

Prof. Dr. Martin Erdmann, 30-Apr-24
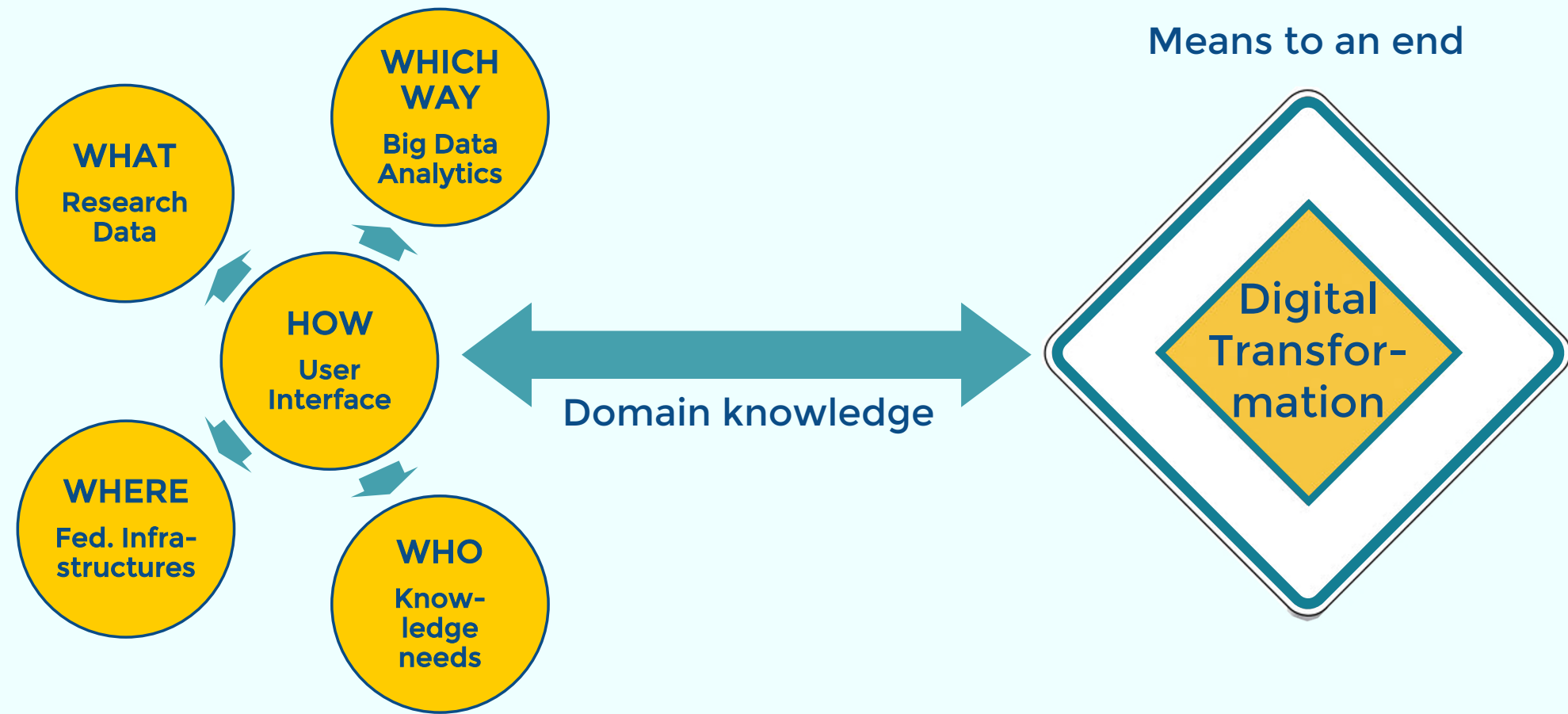
Promote digital transformation in physics research

| Digital Competencies | Networking | Transfer & Communication |

8 research communities
with 20,000 scientists

Community
KAT   Astroparticle
KET   Particle
KfB   Accelerators
KFN   Research with Neutrons
KFS   Resarch with Synchrotron Radiation
KFSI   Research with nuclear Probes and Ions
KHuK   Hadrons and Nuclei
RDS   Astronomy

Ministry science & education
funding program „ErUM-Data"

120 M€ / 2021-2030 for project funding

Network & Transfer Office
ErUM-Data-Hub
2,5M€ / 4 years

Martin Erdmann, RWTH Aachen University

DIG-UM

Collaboration-type Community Organization DIG-UM

TOPIC GROUPS

A particular challenge for the DIG-UM consortium representing research communities with altogether 10,000 researchers is the heterogeneity of the working groups from the different

OVERVIEW BOARD

The OB has directive responsibility for the DIG-UM consortium. Members of the OB are the elected chairpersons of the 8 ErUM committees, and the chairperson of the Resource Provider

DIG-UM

DIGITAL TRANSFORMATION IN THE RESEARCH ON UNIVERSE AND MATTER

GUIDELINES

CONTEXT

Research in the field of Universe and Matter brings together about 10,000 scientists from German universities and research centers working on astrophysics, astroparticle physics, hadron and nuclear physics, particle physics, accelerator research, as well as research with photons, neutrons and ion beams. What unites these scientists is research at leading national and international research facilities such as accelerators with their experiments and observatories. Big data and the ability to manage the upcoming data avalanche will be a prerequisite for the scientific exploitation of the next generation of research facilities, and thus for transformative or even revolutionary new scientific discoveries.
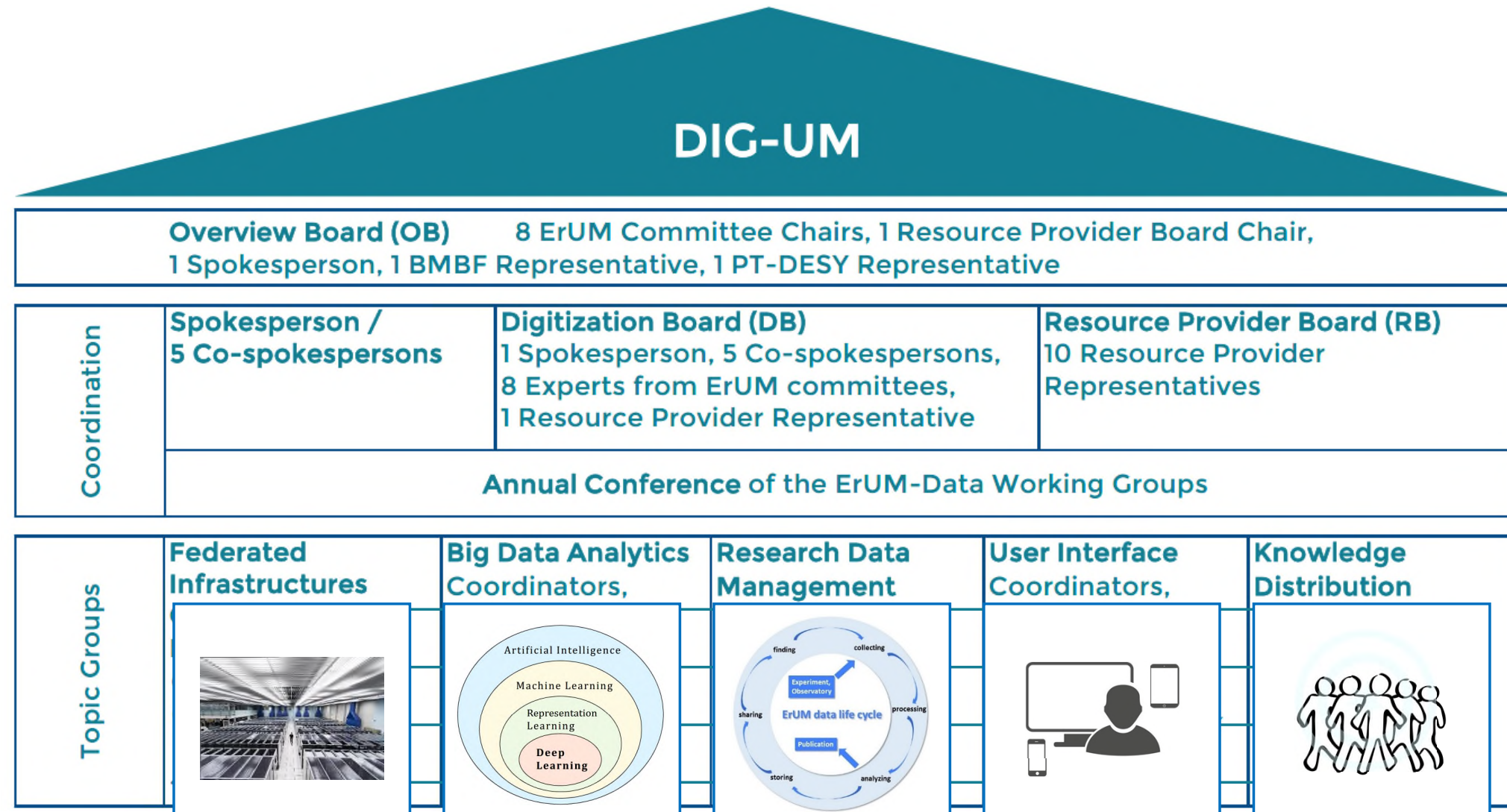
The Federal Ministry of Education and Research (*BMBF*) supports this basic research under the acronym *ErUM* (=Erforschung von Universum und Materie) and has published the Action Plan *ErUM-Data* for funding digital transformation in basic research [1]. A networking and transfer office, referred to as the *ErUM-Data-Hub*, will support activities aiming at the digital transformation process.

DIG-UM ORGANIZATIONAL STRUCTURE

With the foundation of *DIG-UM*, the *ErUM* scientists constitute a bottom-up organization for synergetic research on the digital transformation. The goal of its collaborative work is the joint implementation of the objectives stated in the community white paper *Challenges and Opportunities of Digital Transformation in Fundamental Research on Universe and Matter - Recommendations of the ErUM Committees* [2]. With the here-described *DIG-UM* guidelines, the organizational basis for the collaborative work is laid down. The *DIG-UM* structure is presented in figure 1 and is detailed below. Note that there will be a strong link to the support provided by the above-mentioned *ErUM-Data-Hub*.

19.9.2021

September 19, 2021   1

DIG-UM

Overview Board (OB)	8 ErUM Committee Chairs, 1 Resource Provider Board Chair, 1 Spokesperson, 1 BMBF Representative, 1 PT-DESY Representative	

Coordination	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	Big Data Analytics Coordinators,	Research Data Management	User Interface Coordinators,	Knowledge Distribution

Martin Erdmann, RWTH Aachen University

ErUM-Data-Hub @ Aachen

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

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

Contact: info@erumdatahub.de

erumdatahub.de

Networking

Education

Deep Learning School

Transfer

Hannover Messe

Universum & Materie erforschen.
Potentiale entfalten.
Gemeinsam.

ERUM DATA HUB

DIG-UM

Support, strengthen, expand, boost: ErUM-Data-Hub scope of responsibility

1.
Networking inside

Synergy identification & networking in our communities, funded ErUM-Data projects.

2.
Exchange w outside

Knowledge & innovation transfer with related disciplines, technology, business

3.
Digital competencies

Activities of education & training: schools, workshops, career

4.
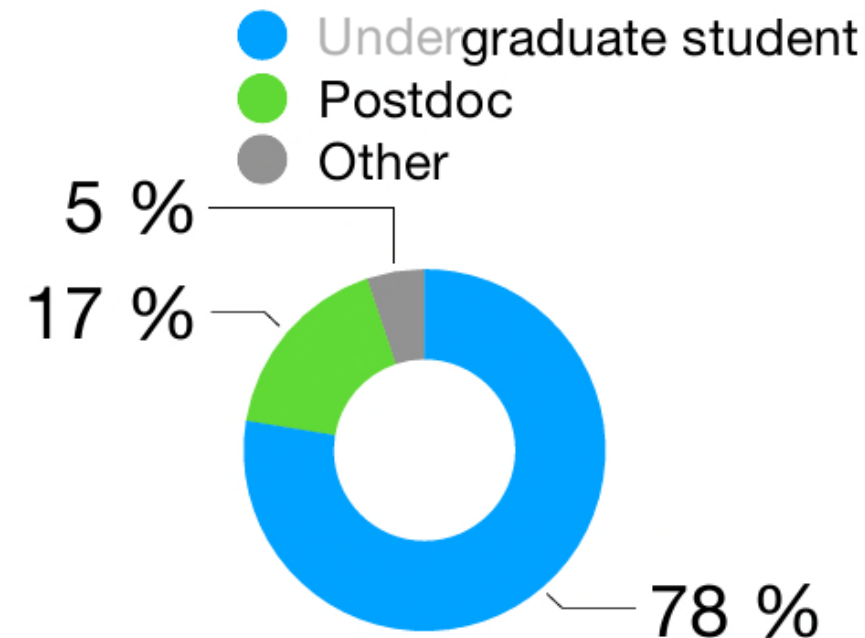## Communication w outside

Public relations & target group-oriented communication of research results

DIG-UM

Deep Learning Schools

Advanced / Basics

- Audience: (PhD) Students

- 4-5 days @hotel

- ~30-50 participants

- 4 Lectures + Tutorials (90')

- Group Challenge

- 3x year

- Undergraduate student
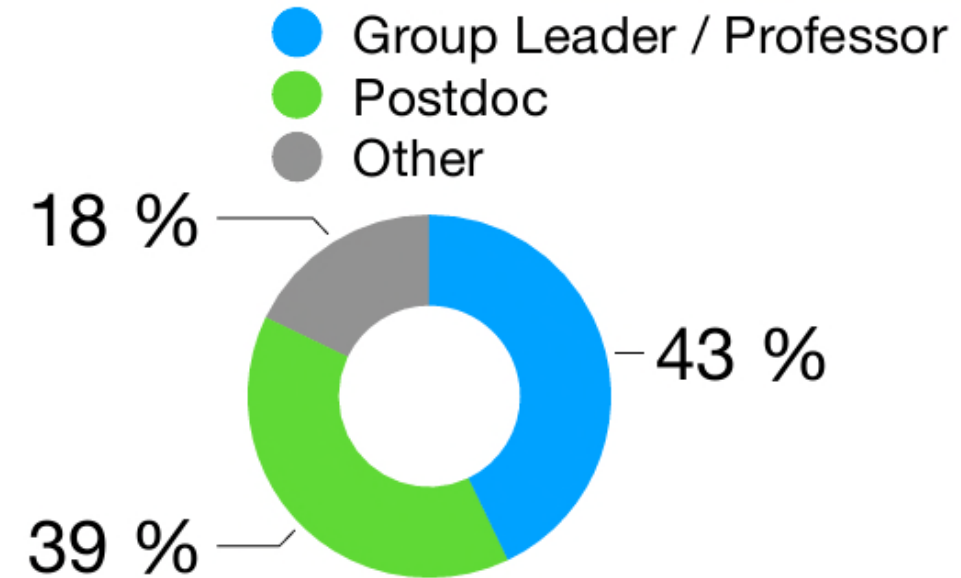- Postdoc
- Other

5 %
17 %
78 %

→ training students to **use** Deep Learning

DIG-UM

Train-the-Trainer

Advanced / Basics

- Audience: Post-docs, Professors, ...

- 2 days @university

- ~30 participants

- 4 Lectures + Tutorials

- 2x year

Group Leader / Professor
Postdoc
Other

18 %

39 %

43 %

→ training lecturers to *teach* Deep Learning

DIG-UM

Teaching digital competencies in carefully selected environments

Event	Location
Train-the-Trainer Deep Learning "Basic Concepts"	RWTH Aachen University
Train-the-Trainer Deep Learning "Advanced Concepts"	Bergische Universität Wuppertal
Deep Learning School „Basic Concepts"	Landhaus Nordhelle, Meinerzhagen
Conceptual Advances in Deep Learning	Hotel zur Post, Wiehl
Active Training Course „Advanced Deep Learning"	Landhaus Nordhelle, Meinerzhagen

RWTH Aachen University

Hotel zur Post, Wiehl

Landhaus Nordhelle, Meinerzhagen

Bergische Universität Wuppertal

DIG-UM

ErUM-Data-Hub Events: Holistic approach on Knowledge Transfer & Networking & Synergies

Disseminating Digital Skills

Enabling Networking

Strengthening communi-cation

ErUM-Data-Hub Events

Identifying synergy effects

Identifying needs

Intensifying exchange

Martin Erdmann, RWTH Aachen University

DIG-UM
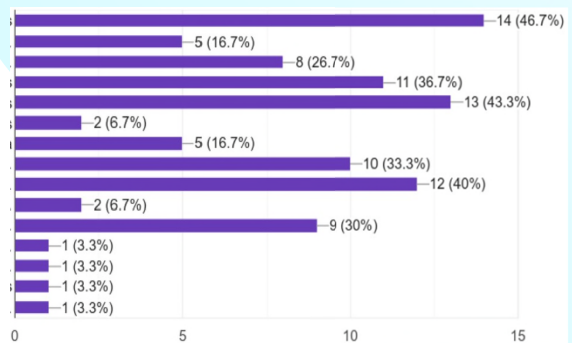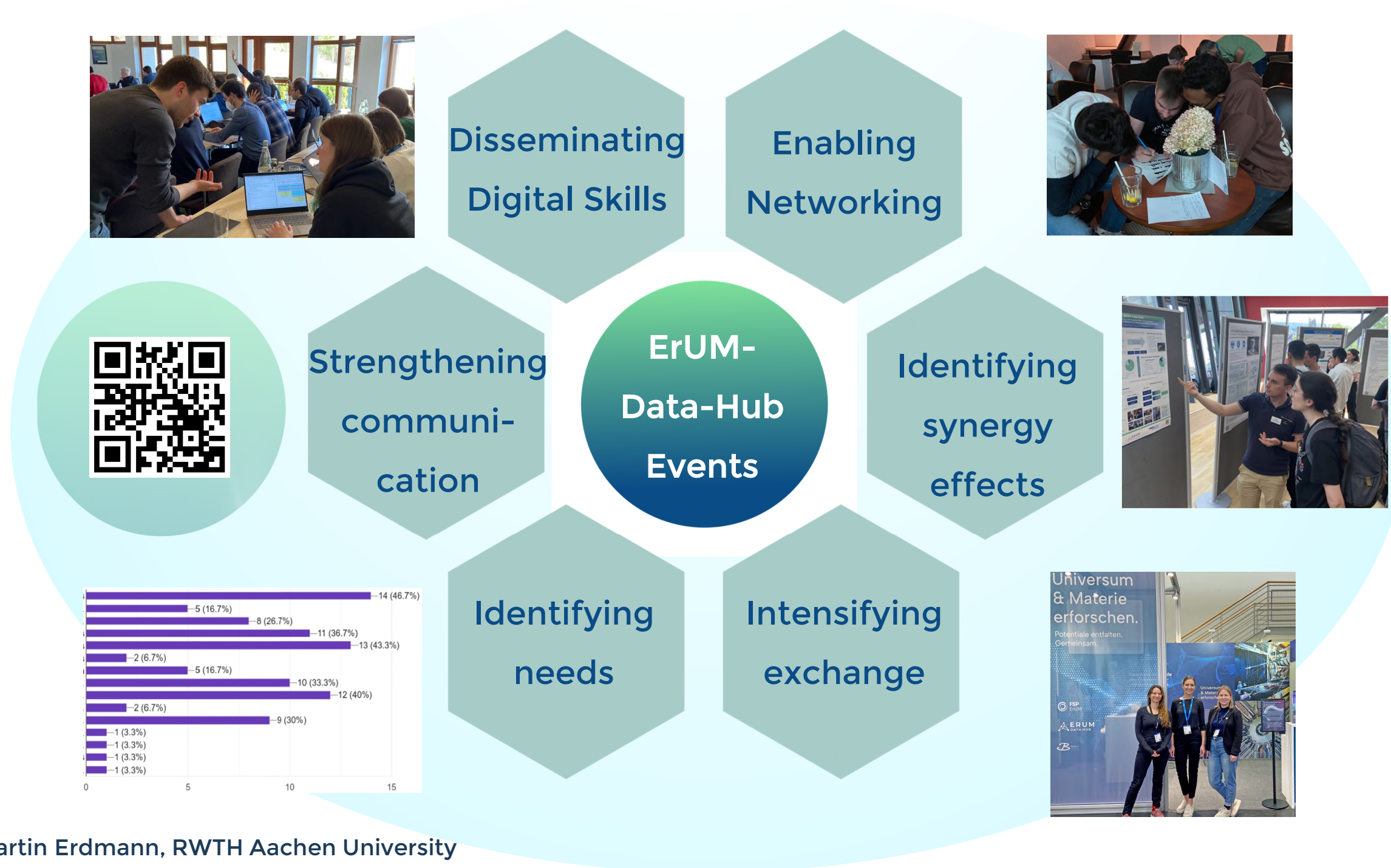
Disseminating digital competencies to a broad ErUM audience

2022

1. Deep Learning Train-the-Trainer Workshop „Basic Concepts"

2. Deep Learning Train-the-Trainer Workshop „Advanced Concepts"

3. Deep Learning School „Basic Concepts"

4. Active Training Course „Advanced Deep Learning"

5. Conceptual Advances in Deep Learning for Research in ErUM
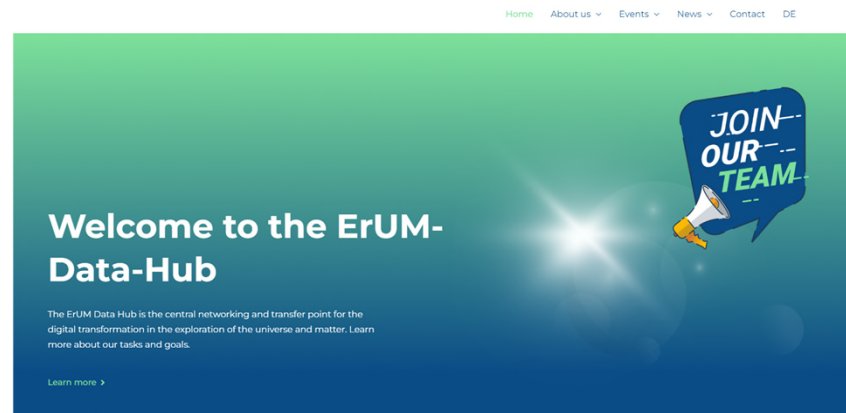
Aachen, Wuppertal, Meinerzhagen, Wiehl

2023

1. **Deep Learning Train-the-Trainer Workshop „Basic Concepts"**

2. **Deep Learning Train-the-Trainer Workshop „Advanced Concepts"**

3. Deep Learning School „Basic Concepts" - Winter School

4. Deep Learning School „Basic Concepts" - Summer School

5. Active Training Course „Advanced Deep Learning"

6. **ErUM-Data Sustainability Workshop (publication)**

7. **ErUM-Data Workshop on Inverse Problems**

8. Research Data Management (DIG-UM Workshop)

9. Big Data Analytics (DIG-UM Workshop)

10. Next Generation Environment for Interoperable Data Analysis (DIG-UM Workshop)

München, Dortmund, Meinerzhagen, Bonn, Berlin, Hamburg

Similar events running in 2024 including
- Fast and Efficient Python Computing School
- Iris-Hep Developers Workshop

Martin Erdmann, RWTH Aachen University

www.erumdatahub.de

DIG-UM

Providing high-quality contents for ErUM Scientists

Launed in February 2024:
The ErUM-Data-Hub Podcast

Welcome to the ErUM-Data-Hub

The ErUM Data Hub is the central networking and transfer point for the digital transformation in the exploration of the universe and matter. Learn more about our tasks and goals.

Learn more >

erumdatahub.de

Aimed at ErUM communities and the interested public

- General Information
- ErUM, DIG-UM
- Events & Documentation
- Contact, Links & Partner
- **Videomaterials**

Deep Learning Train-the-Trainer Workshop "Advanced Concepts"

Am 19. und 20. Juni 2023 luden DIG-UM und der ErUM-Data-Hub zum zweiten Mal zum Deep Learning Train-the-Trainer Workshop "Advanced Concepts" ein. Der diesjährige Workshop fand an der TU Dortmund statt und beinhaltete 4 spannende Voträge von einem herausragenden Speaker*innen-Team zu den Themen Graph Neural Networks (*Rasmus Ørsøe, TUM*), Autoencoder (*Thorben Finke, RWTH Aachen University*), Normalizing Flows (*Dr. Nicole Hartman, TUM*) und Diffusion Models (*Dr. Jakob Knollmüller, TUM*).

Indico
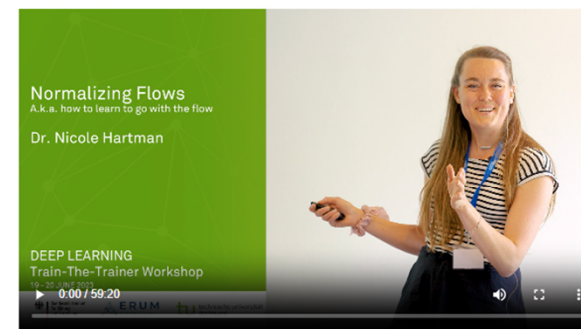
01.

Normalizing Flows

Lecture by Dr. Nicole Hartman

Slides >
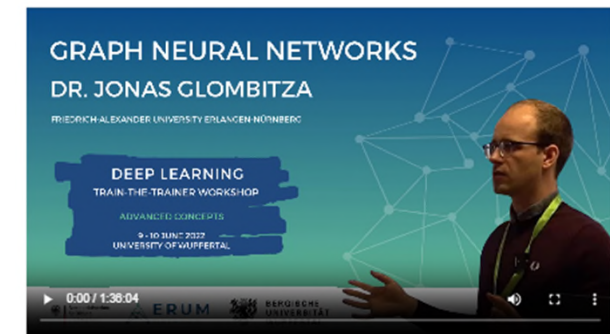
02.

Autoencoder

Lecture by Thorben Finke.

Slides >

More than 500 physicists from all ErUM communities reached

Aachen,
Deep Learning Train-the-Trainer Basic Concepts

Wuppertal,
Deep Learning Train-the-Trainer Advanced Concepts

Wiehl,
Conceptual Advances in Deep Learning
for Research on Universe and Matter

Meinerzhagen,
Deep Learning School „Basic Concepts"

Meinerzhagen,
Active Training Course Advanced Deep Learning
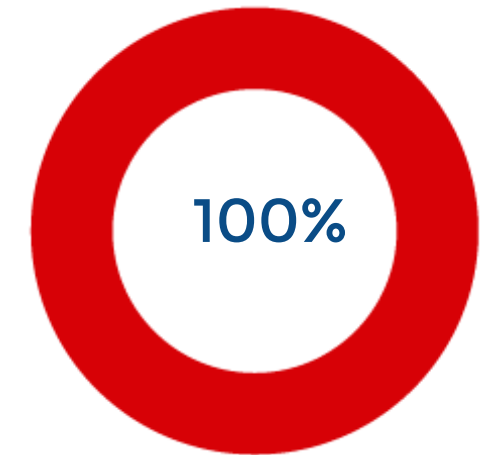
100% Yes

Would you recommend the workshop to colleagues who want to learn basic concepts of Deep Learning themselves?

DIG-UM

Transfer internal – external

- Transfer knowledge between communities

- Isolate issues that are important to communities

Transfer

ErUM-internal

- How do next-gen analyses environments look like?

- How do experiments handle (big) data?

ErUM ↔ Industry

- How to do sustainable computing?

- How is AI used in Industry?

Martin Erdmann, RWTH Aachen University

DIG-UM

Outreach and Transfer

Outreach

- Web presence: erumdatahub.de

- Established network between outreach responsibles

- Lively discussions with broad public Hannover Fair

- We're active on:

  - erumdatahub.de/blog (Blog)

  - @ErUMDataHub

  - @ErUM-Data-Hub

DIG-UM

Sustainability in ErUM Data

4-day Workshop May 23

Developing a culture for sustainable science is growing consensus in community
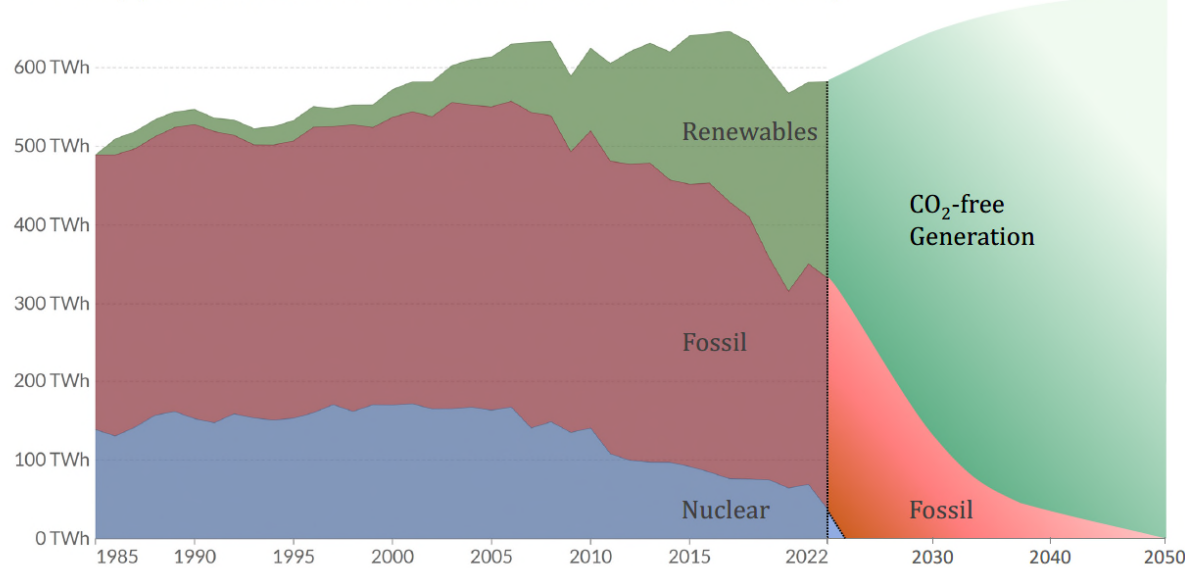
→ Awareness in daily research: balance *knowledge gain* with *resource usage*

https://arxiv.org/abs/2311.01169

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

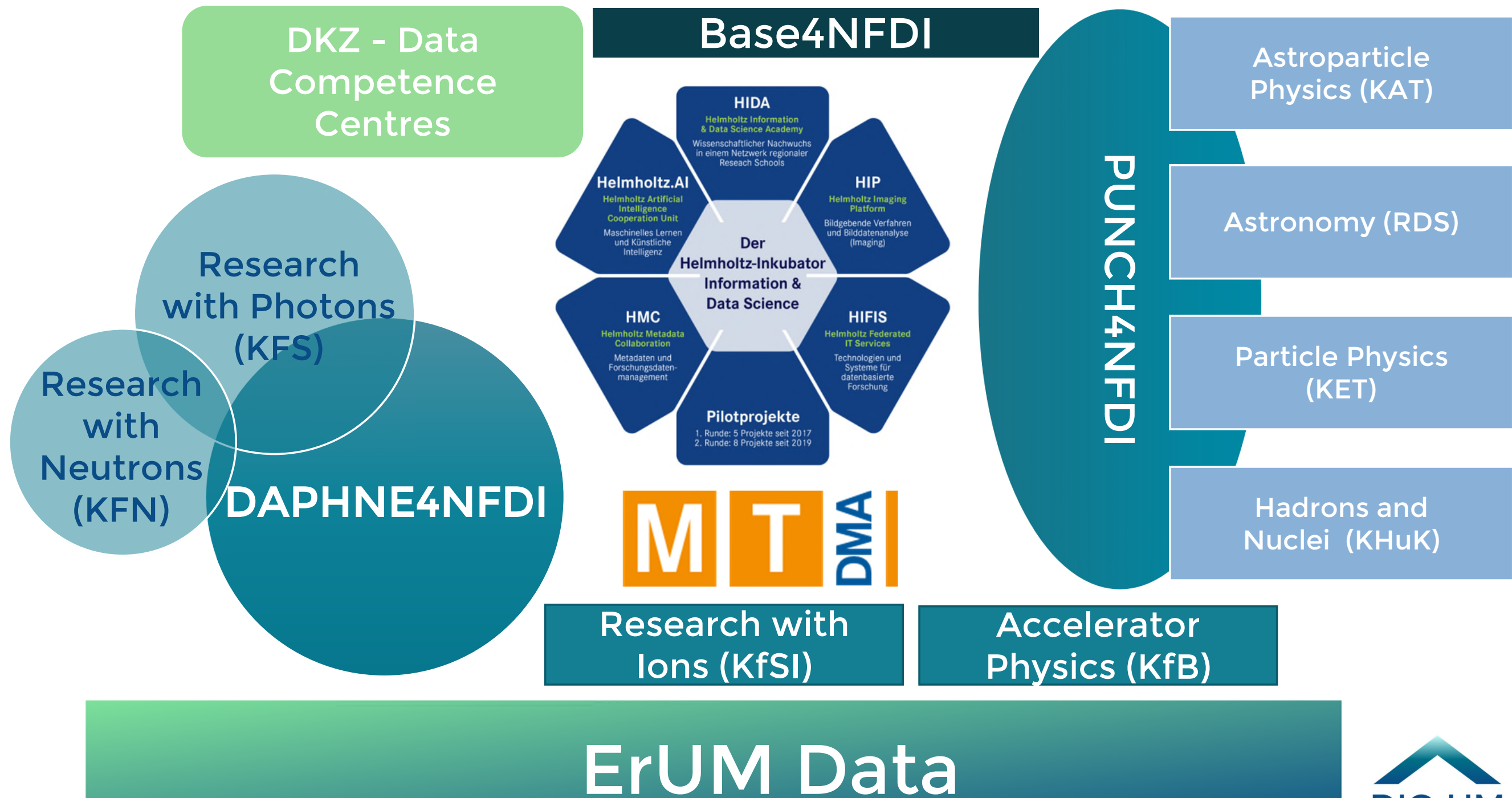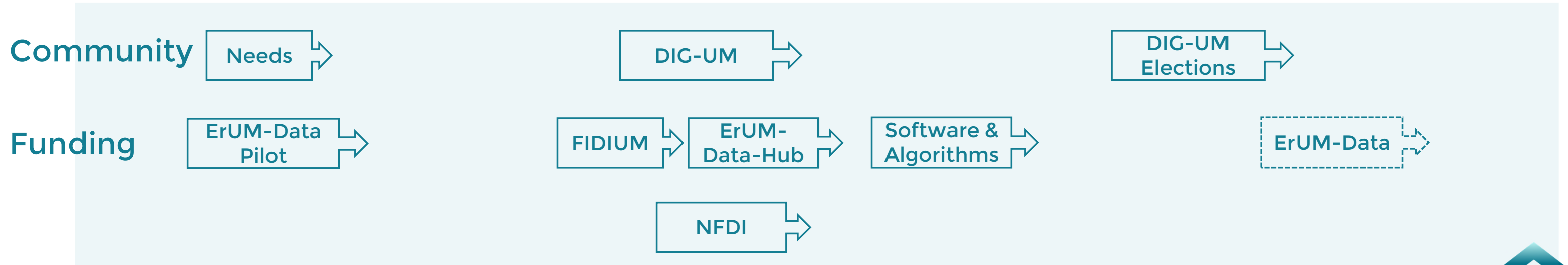## Portfolio of measures

Short term

Medium term

Longer term

Item	Call-to-action
	Immediately or on **short time scale** 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 **medium time scale** 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_2e (water, material etc.).
M10	Train scientists in good practices.
M11	Regularly review and update the CO_2e reduction plan.
M12	Strive to become a role model at all levels and help to establish sustainability in everyday life.
	A **longer term** 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.

Martin Erdmann, RWTH Aachen University

DIG-UM

Multiple Connections on Digital Transformation

International

DKZ - Data Competence Centres

Base4NFDI

Research with Photons (KFS)

Research with Neutrons (KFN)

DAPHNE4NFDI

HIDA
Helmholtz Information & Data Science Academy

Wissenschaftlicher Nachwuchs in einem Netzwerk regionaler Reseach Schools

Helmholtz.AI
Helmholtz Artificial Intelligence Cooperation Unit

Maschinelles Lernen und Künstliche Intelligenz

HIP
Helmholtz Imaging Platform

Bildgebende Verfahren und Bilddatenanalyse (Imaging)

Der Helmholtz-Inkubator Information & Data Science

HMC
Helmholtz Metadata Collaboration

Metadaten und Forschungsdaten-management

HIFIS
Helmholtz Federated IT Services

Technologien und Systeme für datenbasierte Forschung

Pilotprojekte
1. Runde: 5 Projekte seit 2017
2. Runde: 8 Projekte seit 2019

MT DMA

Research with Ions (KfSI)

Accelerator Physics (KfB)

PUNCH4NFDI

Astroparticle Physics (KAT)

Astronomy (RDS)

Particle Physics (KET)

Hadrons and Nuclei (KHuK)

ErUM Data

Martin Erdmann, RWTH Aachen University

DIG-UM

Time Line Digital Transformation

| BMBF Workshop in Bonn | Challenges and Opportunities of Digital Transformation in Fundamental Research on Universe and Matter | Aktionsplan ErUM-Data | DIG-UM | 5 Topic Groups | | ErUM DATA HUB | ErUM-Data-Hub Phase 2 |

2018 → **2019** → **2020** → **2021** → **2022** → **2023** → **2024** → **2025** → **2026** → **...**

Community
- Needs
- DIG-UM
- DIG-UM Elections

Funding
- ErUM-Data Pilot
- FIDIUM → ErUM-Data-Hub → Software & Algorithms
- NFDI
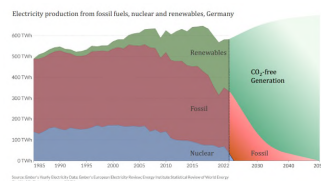- ErUM-Data

Summary
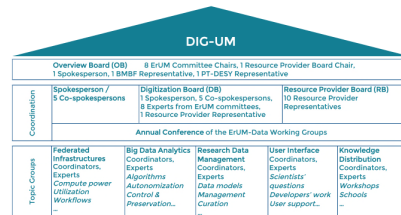
- **Newly built large Community: well alive & very active.**

- **Major challenges ahead:**
  - ➢ **We must prepare for a future with exponentially increasing data volumes, without exponentially increasing resources.**
  - ➢ **Exploiting the upcoming science data constitutes major compute challenges to us (hardware, services, algorithms).**
  - ➢ **We need to push our digital transformation strongly to reach up to US and others - has impact far beyond our research.**
  - ➢ **Sustainability measures constitute major challenges.**

Martin Erdmann, RWTH Aachen University

DIG-UM

Backup

Martin Erdmann, RWTH Aachen University

DIG-UM
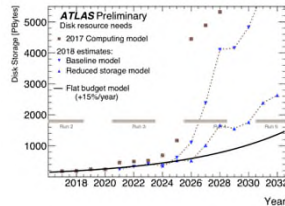
Connection to ErUM-Data Themes

- Components for data management
- Metadata administration & access
- Data Workflows experiments
- Data life-cycle management
- Data Access rights

- Sustainabe software
- Modern technologies
- Efficient use of data
- Maximize scientific gain
- Respecting resources

WHAT
Research
Data

WHICH WAY
Big Data
Analytics

ErUM-Data

| Forschungsdatenmanagement |
| Software & Algorithmen |
| föderierte Digitalinfrastrukturen |

HOW
User
Interface

- Sciencists' exploratory interface
- Self-explanatory design
- High-quality data products
- Worksflows
- Algorithms

- Data Lake
- ErUM-Federation
- Authentication
- Blind Discovery
- Sustainability

WHERE
Federated
Infrastruct.

WHO
Knowledge
need

- Embedding gained Knowledge
- Travel money workshops/school
- Tailored learning projects

Martin Erdmann, RWTH Aachen University

DIG-UM