Zoommeeting Digitization Board 9.3.2023, 8:30-9:25

KFN, Tobias Richter, KAT, Andreas Haungs, KFS, KFSI, KfB, RDS, KHuK, KET, FI, Markus Demleitner, BDA, Thomas Kuhr, Jan Steinheimer, RDM, Astrid Schneidewind, UI, KD, Dirk Lützenkirchen-Hecht, Judith Reindl, RPB, Günter Duckeck, ErUM-Data-Hub, Peter Fackeldey, Benjamin Fischer, SP, Martin Erdmann

Big-Data-Analytics Connection Workshop in Hamburg 23./24.2.23
The workshop by Gregor Kasieczka, Thomas Kuhr, Jan Steinheimer: The workshop brought the DIG-UM structure to life, many people got to know each other for the first time. 8/10 funded consortia presented their project, 1 non-funded consortium as well. Data analyses in NFDI were also presented. Grants have primarily been awarded to consortia with cross-community partners and industry partners. An annual meeting of the consortia is targeted, and consortium leaders are on the Big Data Analytics Strategy Group mailing list.

Deep Learning Basic Concepts School, Meinerzhagen 27.2.-2.3.23
The school was carried out onsite by Angela Warkentin and Benjamin Fischer: In addition to the established program of Neural Network Building Blocks, Mastering Model Building, Convolutional Neural Networks, Introspection of Neural Networks, the school had a Challenge that was worked on in small groups. The school was again very well received and a great success overall.

ErUM-Data – NFDI Workshop in Bonn 13./14.3.23
The rich workshop agenda is completed (Monica Valencia-Schneider, Astrid Schneidewind, Kilian Schwarz). It will emphasize the continuous developments, the overlap between ErUM-Data plus physics-related NFDI consortia plus Helmholtz DMA, introduce envisaged Data Competence Centers, present concepts of large-scale data storage, and will include several small working groups. Overall, the goal is a white paper to achieve concrete input for the strategy meeting for the next funding period.

Sustainability Workshop 30.5.-2.6.23
The workshop goal is to produce a short report on strategic concepts for sustainability transformation in ErUM-Data from the funding applications on. The 12 leading questions (*appended below*) have been formulated, the program of the keynote speakers is currently being designed by Michael Bußmann, Markus Demleitner, Günter Duckeck, Martin Erdmann, Peter Fackeldey, Benjamin Fischer, Stefan Funk, Thomas Kuhr, Michael Lupberger, Pardis Niknejadi, Markus Roth, Astrid Schneidewind, Achim Streit, Angela Warkentin.

Update & perspectives on ErUM-Data-Hub Team
The selection process for the position on communication is ongoing (two applicants).
We have capacities for additional Referent:innen and have initiated a further call:
https://www.rwth-aachen.de/go/id/kbag/file/V000004777
New is that `external´ Referent:innen can be considered.
Everybody is asked to look for suitable persons and talk to Martin Erdmann.

Request for a Meeting of the DIG-UM Overview Board

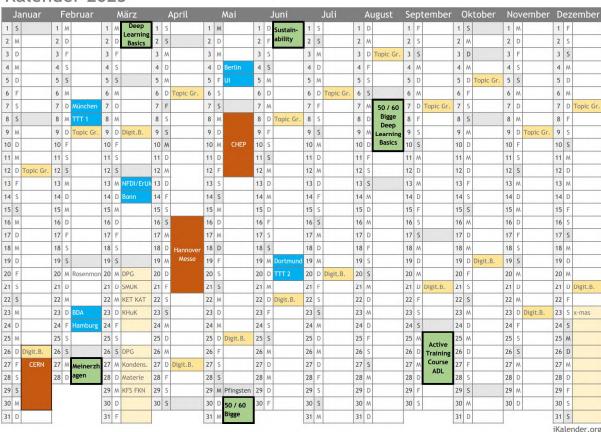We recommend a meeting of the overview board with the following topics:

− Election of the OB-Deputy
− DIG-UM: Report of the Spokesperson on achievements and challenges
− ErUM-Data-Hub: Report on achievements in 2022 compared to the original application
− ErUM-Data-Hub: Envisaged Programme 23/24 (will be fixed on 16.3., 16:00)
− Sustainability in ErUM-Data: PRISMA-Trialog and Workshop → White Paper
− Publications in DIG-UM: authorship and relation of authors to the ErUM committees (White Papers, Journal-Publications, DIG-UM Denkschrift (in 3 years from now)).
− Aob

For the authorship discussion, it needs a clever suggestion how to solve it such that the overview board will come to a decision within this meeting. Initial ideas are a footnote, acknowledgement, author (member of the ErUM committee), citation of a survey etc.

Future monthly meetings
- Next Meeting 6-Apr-2023, 8:30-9:30 **Topic Groups**
- Next Meeting 27-Apr-2023, 8:30-9:30 **Digitization Board**

Kalender 2023

https://indico.desy.de/event/37480

We agreed on 12 leading questions to the 6 working groups that will drive discussions & write-up:

Wednesday: Hardware & Research Data

1. **Footprint:** Construct a comprehensive mental picture of the footprint of all ErUM-Data related activities. Where does quantitative knowledge exist or is lacking? What resource needs do you see, what opportunities for savings? What innovations are needed to keep sustainable use of resources in balance with demands? To what extent does continuing education play a role? How feedback can reduce a footprint through ML methods.

2. **(Dynamic) Energy Supply:** Where to locate & operate computing systems incl. storage. How could a dynamic energy supply look like, which covers the needs of ErUM-Data related activities with renewable energies to a large extent? What information flows would be required for this? What mechanisms and what dynamics are required on a supra-regional basis to create compensation possibilities for windless/sunless periods?

3. **Hardware Lifetime:** How could prolonged/optimized usage of hardware resources in view of technology evolution be modeled beyond their usual lifetime? What short- and medium-term monitoring would be required to signal indispensable replacements on the one hand, and to execute computing jobs matching their algorithmic requirements on prolonged or current hardware on the other?

4. **Hardware & Algorithms:** Which adaptive measures for hardware and algorithms could have a decisive impact on ErUM-Data? Which types of hardware (including e.g. GPU, TPU, FPGA, neuromorphic computing) could be considered, which automated mechanisms exist for adapting algorithms to non-specific or dedicated hardware?

5. **Smart Data:** Deciding when and how to discard information without losing scientific value, based on learning from nature and experiment. What mechanisms for transforming data to smart data can be envisioned, and how can evaluation and control of information gain or loss be accomplished. How can archiving and retrieving data be managed?

6. **Cultural Change:** How could a comprehensive educational area for rethinking, among other things, the use of computer hardware, actually required information (smart data), preparation of data packages (event loops versus event chunks), etc. look like. How can we change to a culture of data reuse? Assess ethical implications and risk assessment.

Thursday: Algorithms & Mindset

1. **Autonomization:** We witness the transformation from the era of automation to an era of autonomization (e.g. unsupervised learning). Where will ErUM-Data benefit from autonomization, which innovations are necessary and how can the reliability of the autonomously obtained results be ensured?

2. **Inquiries & Dynamics:** How can input questions be optimally posed to generate the best possible output from the machines? What relevance will dynamic learning algorithms and machines have for the field of ErUM-Data?

3. **Algorithmic & Software:** Our thinking in algorithms and software has a direct impact on resource requirements. How can a sustainable algorithm & software engineering and an associated educational program in algorithm & software development look like to get ErUM-Data in the forefront of developers?

4. **Machine Models:** Pre-trained and generative models have high potential for energy savings in both their creation and usage of machine learning. What innovations are needed to achieve reliable routine operation?

5. **Injected Intelligence:** How can reasoning by the physicist, mathematician, or any other kind of intelligence speed up the processes of learning or make them more energy efficient? What measures can we use to avoid constantly reinventing the wheel? How can knowledge discovery of work already performed look like?

6. **Workflow & Stakeholders:** How can well-defined, reproducible workflows with high user dynamics (data analyses) be captured that remain functional in the long term? How can an overall picture be created with all stakeholders working together on a large-scale project for the benefit of sustainability across their departmental boundaries?

Information in Wiki: https://wiki.erumdatahub.de/de/Sustainability