# Big Data Analytics Report 2024

DIG-UM Annual Meeting 2024

Jan Steinheimer and Thomas Kuhr

DIG-UM

# Almost 2 years into first funding round – Big topics for the next round

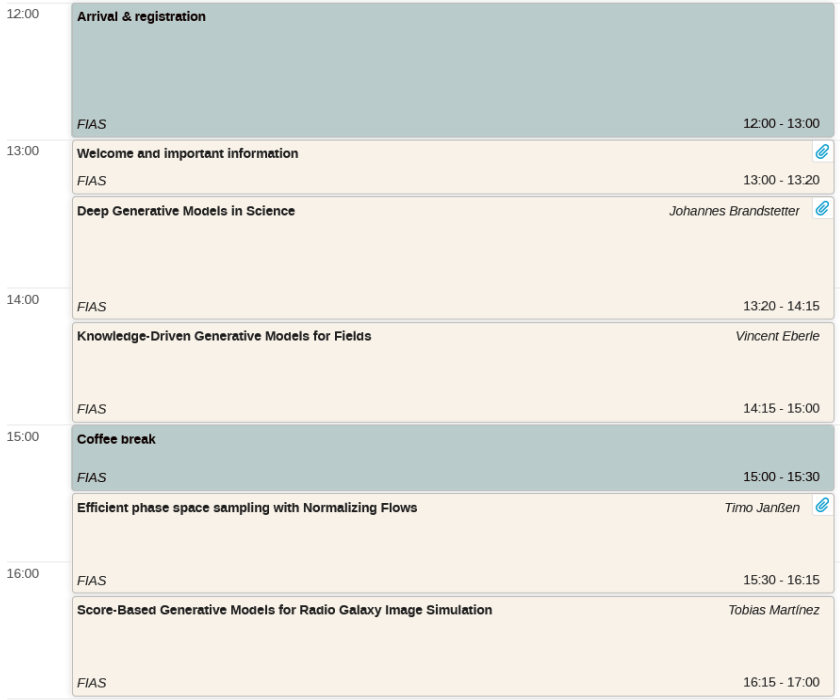
- We have 10 consortia which are running.
- Participation of all ErUM-communities.
- The majority of consortia at least planned to submit a new proposal.

Based on workshops, TG meetings ((Bi-)Monthly TG ~ 10 participants.)
and feedback on planned consortia:

- AI on dedicated hardware (e.g. FPGA) for fast and big data.
- Generative models.
- Sustainable and efficient software and algorithms.
- Large Language Models and how they can transform scientific work.

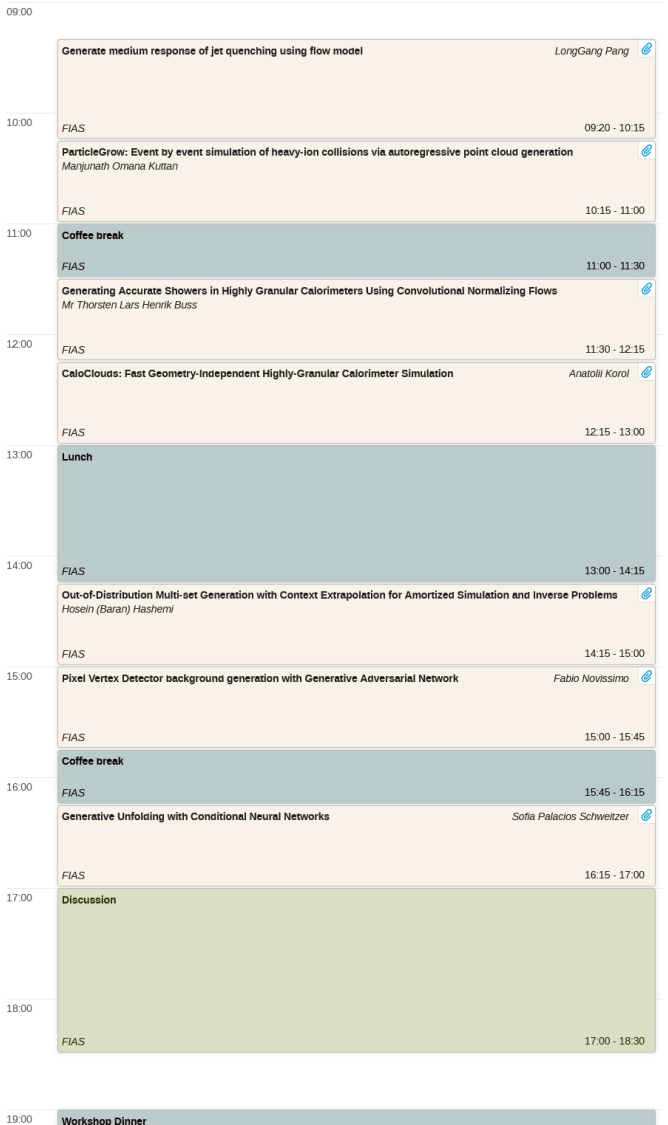
# Workshops 2024 (supported by Hub and TG)

- HuB and TG supported workshops this year
- March: Generative Models at FIAS (Frankfurt)
- Contributions from various fields – growing importance



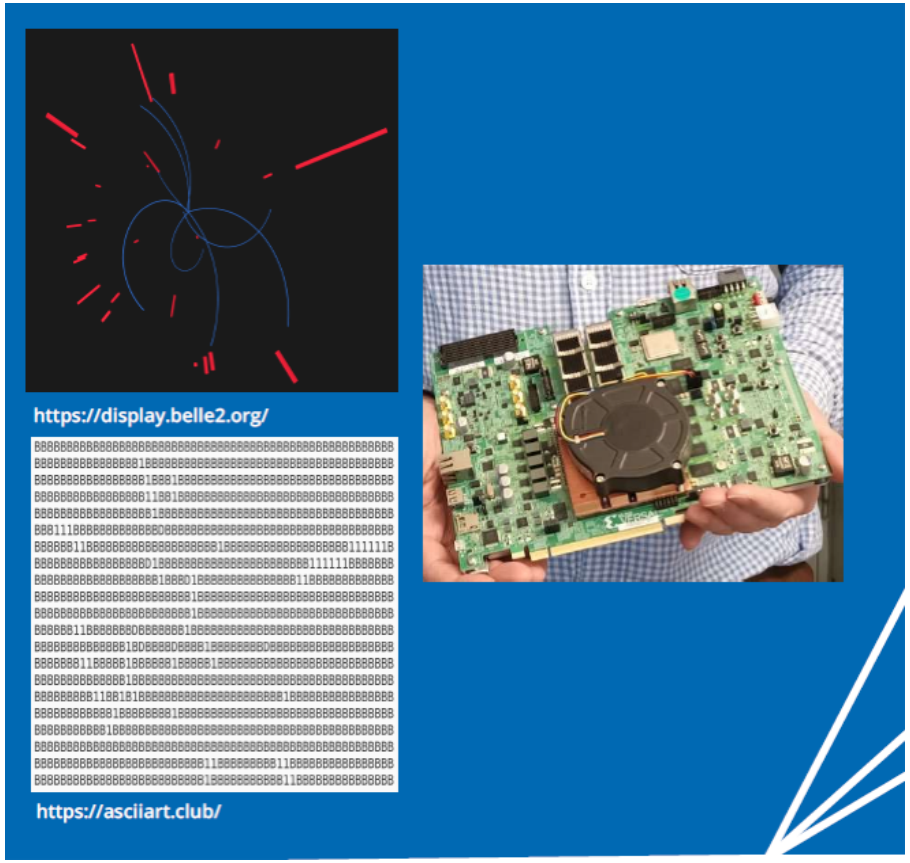

**Quo vadis science?**

- Large data era is hitting scientific research
  -> Foundation models for science
- What is multi-modality in physics?
  - E.g. Eulerian vs Lagrangian in CFD
- Does it always have to be the data? Can't it be constraints?
  - In contrast to text, images, videos, we can write down constraints in science
- Diffusion principle is destined to stay (very personal point of view)



# Workshops 2024 (supported by Hub and TG)

- April 8–11 : Workshop on Realtime Machine Learning (Giessen)
- Large interest in DL on FPGA for ‘realtime’ analysis.



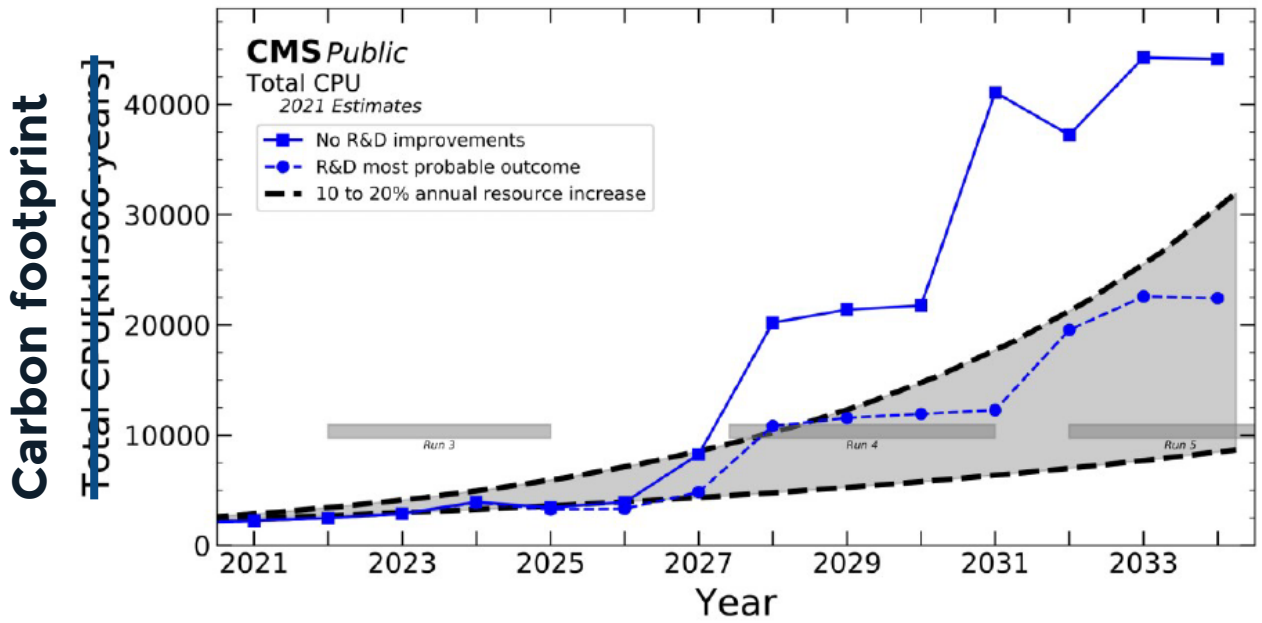

## Closing comments (Sören Lange):

- Triggerless DAQ systems is a modern trend (4 experiments reported at this workshop);
- Sending data with Terabytes/s (CBM ~1 TB/s, ALICE >3 TB/s);
  pushing all data to CPU farm is tempting (“easier” than FPGA programming)
- System-on-a-chip (i.e. FPGA interfaced with ARM processors) changed the FPGA world, makes it much easier to get e.g. NN training data to the FPGA (compared to “pure” FPGAs); non-expensive boards available (ZYNQ) and AMD/Xilinx kept the approach for newer (and more expensive) platforms
- Surprising ideas presented:
  - “half autoencoder” (CMS), score is taken from the bottleneck instead of loss function from behind the decoder
  - anomaly detection using decision trees instead of autoencoder;
  very tempting, does not need multiplications (DSP slices) but only “if statements”
- Versal is the new Porsche

IG-UM

# Sustainable research

- Already 2023 the issue was published.
- The next round of funding may take the suggestions into account.




Resource-aware Research on Universe and Matter:
Call-to-Action in Digital Transformation

Ben Bruers[1], Marilyn Cruces[2], Markus Demleitner[3], Guenter Duckeck[4], Michael Düren[5], Niclas Eich[6], Torsten Enßlin[7], Johannes Erdmann[6], Martin Erdmann[6*], Peter Fackeldey[6], Christian Felder[8], Benjamin Fischer[6], Stefan Fröse[9], Stefan Funk[10], Martin Gasthuber[1], Andrew Grimshaw[11], Daniela Hadasch[9,12], Moritz Hannemann[8], Alexander Kappes[2], Raphael Kleinemühl[13], Oleksiy M. Kozlov[14], Thomas Kuhr[4], Michael Lupberger[15], Simon Neuhaus[13], Pardis Niknejadi[1], Judith Reindl[16], Daniel Schindler[17], Astrid Schneidewind[8], Frank Schreiber[18], Markus Schumacher[19], Kilian Schwarz[1], Achim Streit[20], R. Florian von Cube[20], Rodney Walker[4], Cyrus Walther[9], Sebastian Wozniewski[17], Kai Zhou[21]

# Community feedback: Thank you Hub!

- Besides supporting the topical workshops and the annual TG meeting, the Hub also helps to make other BDA related workshops known to the community.
- „Good outreach and impact“

# Next BDA annual meeting

- Save the date: March 13-14 in Munich.
- Thank you: Thomas Kuhr and Christian Stieghorst for helping with the organization.

Contact information:

https://wiki.erumdatahub.de/de/mailing-lists