KHuK Annual Meeting

Preparation of the ErUM-Data Strategy Meeting

7.12.2023 I TOBIAS STOCKMANNS

Responds from KHuK Community

General questions

- Funding for theory ➔ should be no problem
- Funding for Helmholtz centers ➔ easier than in ErUM-Pro
- Size of groups/consortia ➔ request: more flexibility
- Involvement of external (industry) partners ➔ under discussion
- Discrimination to NFDI ➔ complementary
- Funds available / financial means per application?

Responds from KHuK Community

Ideas for projects/topics

- Big Data Analytics:
  - Reproducibility of complex amplitude models
  - Access to particle properties and decay information via JULIA language
  - Nuclear PDF with LHC data
  - Joint inference of calibration and signal
  - AI methods for data analysis
  - Realtime algorithms, ML / ML in DAQ
  - Algorithms for heterogeneous computing on CPU & GPU
  - Generalized tools for particle decay reconstruction (KFParticle)
  - Track reconstruction and alignment

- User Interfaces:
  - Interface for variable data sets of FRS at GSI

Responds from KHuK Community

Ideas for projects/topics

- Federated Infrastructure:
  - Green IT as overarching topic of ErUM-Data: Resource savings by software/hardware optimization
  - Cross-platform framework developments (e.g. XPU project) and applications
  - Energy efficiency of heterogeneous computing on CPU & GPU

- Research Data Management:
  - Open Data handling / workflows
  - Coupling of metadata in online/offline workflows for data-intensive experiments
  - Connection of publications with compute elements
  - Benchmarking of relevant workflows for evaluation of resource efficiency
  - Optimization of resource efficiency of software (used particularly in the KHUK community) for dynamical models of heavy-ion collisions

Topic Groups

in DIG-UM

What should be stored?

What?

Research Data Management

Where should it be stored?

Where?

Federated Infrastructures

How should the data be accessed?

How?

User Interface

Big Data Analytics

Which?

Which algorithms should be used?

Knowledge Distribution

Who?

Who needs to know about this?

Federated Infrastructures

- Networked IT systems with common standards, and shared facilities for discovery, authentication…

- Authentication and Authorization
- Distributed storage too large for one site
- Shared computing resources with access to distributed storage
- Analysis platforms: interactive analysis of large scale data on federated resources
- Archiving following F.A.I.R standards ➔ connection to RDM

Contact person: Markus Demleitner
(msdemlei@ari.uni-heidelberg.de)

Research Data Management

Topics suggested for funding

- Development of (modular) systems for research data management
  - Standardized solutions for the management of heterogeneous scientific data
  - Interfaces and programs for easy access and management of large data samples
  - Interfaces and programs to catalog and find large data samples
- Development of systems to manage and handle meta data
- Development of data workflows to efficiently steer and perform experiments
  - Reduction of raw data - lossless and lossy
  - Processing of data from similar experiments
  - Visualization
- Management of access rights

Contact person: Astrid Schneidewind
(a.schneidewind@fz-juelich.de)

User Interfaces

Topics

- Experiments should be encouraged to publish their data in a timely manner
- Interdisciplinary analysis require cross-community standards
- Accessing data is not enough:
  - Capturing and archiving workflows
  - ➔ Finding, using and combining existing solutions needs to be simplified
- Knowledge needs to be decoupled from individual persons
  - Utilize and organize existing code and workflows
  - Semi-automatic generation of code
  - PhysicsGPT

Contact person: Pierre Schnizer
(pierre.schnizer@helmholtz-berlin.de)

Knowledge Distribution

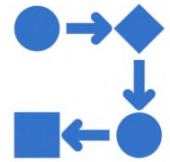
Contact person: Dirk Lützenkirchen-Hecht
(dirklh@uni-wuppertal.de)

Data-Science Themen stärker in den Universitäten verankern

Dauerhafte Verankerung von Inhalten in den Studiengängen

Lehre von Digitalen Kompetenzen

Beginn bereits im Bachelorstudium



Entwicklung von “tailored-learning” Konzepten

Unterschiedliche Konzepte zur Wissenvermittlung unterschiedlicher Themen werden erarbeitet und öffentlich zugänglich gemacht (z.B. Problem-based learning)

Qualifizierung des Nachwuchses

Coaching Angebote

Hilfe zur Selbsthilfe (z.B. Working out loud)

Einbeziehen der breiten Öffentlichkeit

Podcast

Wissensvermittlung

Physik GPT

Verankerung von Knowledge Distribution in den Ausschreibungen

Jedes geförderte Projekt muss Konzepte der Community integrieren oder eigene Konzepte verwirklichen

Big Data Analytics

- See slides from Jan

How to continue

- Stay in touch:
  - ErUM-Data Community: DIG-UM Annual Meeting
    11.12.2023 14:00 – 16:00 https://indico.desy.de/event/42173/
  - Topic Groups (https://erumdatahub.de/dig-um/)
    - **Federated Infrastructures** - Markus Demleitner (msdemlei@ari.uni-heidelberg.de) / Kilian Schwarz (kilian.schwarz@desy.de)
    - **Big Data Analytics** - Thomas Kuhr (Thomas.Kuhr@lmu.de) / Jan Steinheimer (steinheimer@fias.uni-frankfurt.de)
    - **Research Data Management** - Astrid Schneidewind (a.schneidewind@fz-juelich.de) / Monica Valencia-Schneider (vs@uni-koeln.de)
    - **User Interfaces** - Pierre Schnizer (pierre.schnizer@helmholtz-berlin.de) / Tim Ruhe (tim.ruhe@tu-dortmund.de)
    - **Knowledge Distribution** - Dirk Lützenkirchen-Hecht (dirklh@uni-wuppertal.de) / Judith Reindl (judith.reindl@unibw.de)

- KHuK-Computing mailing list: https://www-listserv.gsi.de/cgi-bin/wa?SUBED1=KHUK-COMPUTING&A=1
- Planned: Open KHuK computing meetings