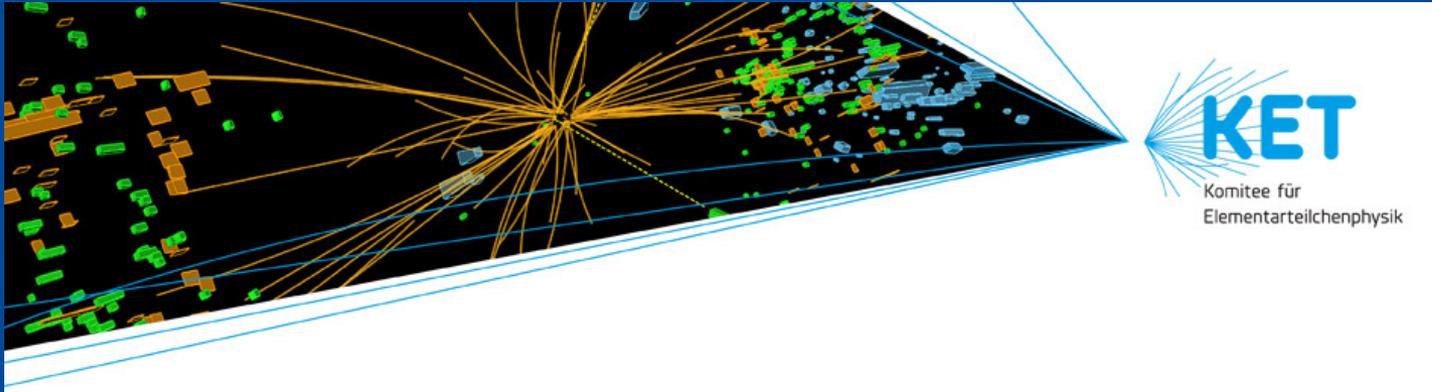


# Introduction and a few remarks



## „Perspektiven für HEP-Software und -Computing in Deutschland“

Albert-Ludwigs-Universität Freiburg



**UNI  
FREIBURG**

Markus Schumacher

KET-Computing&Software-Workshop

Wuppertal, 27.9. 2018

# Goals of the Workshop



- Get at least partial overview of computing activities in KET community
- Discussion of computing models for HL-LHC and others experiments (usage/distribution of hardware and mode of operation, software developments,...)
- **Provide a consensus input for the BMBF Workshop on „ErUM-Data“**
- Discuss options and plans for participation in NFDI and research data management in general
- ....  
(list not exhaustive, but we only have 1 day)

44 registered participants (today 9:00am) including fortunately a few younger ones

# BMBF-Workshop „ErUM-Data“



- Location: BMBF Bonn Date 4./5. October
- Participation upon invitation from BMBF (approx 120 participants)
- From KET-Community:  
V. Büscher (deputy KET), M. Elsing (KET member),  
T. Kuhr (spokesperson of compound IDTErUM)  
G. Quast A. Schmidt, M. Schumacher (Prisma-Trialog) ,  
nominated by KET chair  
plus K. Desch, E. Elsen, V. Guelzow, A Streit,
- Goals:
  - Deliver input for BMBF-Aktionsplan „ErUM-Data“
  - Develop further strategy paper of committees KAT, KET, KFB, KFN, KFS, KHUK, RDS bzw.
  - Produce a paper with recommendations until ???
- Plenary sessions plus parallel sessions on key topics:
  - Föderierte Infrastrukturen (Federated Infrastructures)
  - Big Data Analytics
  - Forschungsdatenmanagement (Research data management)

**Prisma-Trialoge:** Zur Weiterentwicklung eines Querschnittsthemas oder Handlungsfeldes werden Trialoge einberufen. Sie führen die Sichtweisen von Fachleuten aller drei Prisma-Seiten zusammen. Das Bundesforschungsministerium lädt dazu Vertreterinnen und Vertreter von Hochschulen, Wissenschaftsorganisationen und internationalen Einrichtungen ein. Sie analysieren das Handlungsfeld oder Querschnittsthema und leiten Handlungsempfehlungen ab. Dazu entwickeln sie abgestimmte Konzepte, die in die Entscheidungsfindung des Bundesforschungsministeriums für neue Fördermaßnahmen einfließen.



# Program of BMBF-Workshop „ErUM-Data“



UNI  
FREIBURG

Donnerstag, 04.10.2018

- 12:00 - 13:00 Registrierung und Imbiss (Foyer vor Saal 2)
- 13:00 - 13:15 Begrüßung, *MinDirig Dr. V. Dietz, BMBF*
- 13:15 - 15:15 Einführung in das Thema (jeweils inkl. 10 Min. Diskussion)
- 13:15 Zielsetzung des Strategieworkshops, *RD'in Dr. A. Fischer, BMBF*
- 13:45 Vorstellung des Strategiepapiers, *Sprecher/in Prisma-Trialog ErUM-Data*
- 14:15 Impulsvortrag „Herausforderungen bei Großexperimenten“, *Prof. Dr. E. Elsen, CERN*
- 14:45 Impulsvortrag „Herausforderung aus der Wirtschaft“, *N. N.*
- 15:15-15:45 Pause (Foyer vor Saal 2)
- 15:45-16:45 Vorstellung Initiativen (jeweils inkl. 10 Min. Diskussion)
- 15:45 Nationale Forschungsdateninfrastruktur, *RfII Sprecher*
- 16:15 Europäische Initiativen: EOSC, FAIR-Data, *Dr. A. Karalopoulos, EC*
- 16:45-17:45 Arbeitsgruppen (AG) Einführung (jeweils inkl. 5 Minuten Diskussion)
- 16:45 Vorstellung AG 1: Föderierte Infrastrukturen, *Leiter der AG*
- 17:05 Vorstellung AG 2: Big Data Analytics, *Leiter der AG*
- 17:25 Vorstellung AG 3: Forschungsdatenmanagement, *Leiter der AG*
- 17:45-17:50 Aufteilung in Arbeitsgruppen
- 17:50-18:20 Impulsvorträge in den einzelnen Arbeitsgruppen (inkl. 10 Minuten Diskussion)
- PGR AG 1: Föderierte Infrastrukturen, *N.N.*
- Saal 2 AG 2: KI, *N.N.*
- Saal 1 AG 3: Datenmanagement, *N.N.*
- 19:00-21:00 Abendessen mit Möglichkeit zur Vernetzung und Diskussion an Thementischen (Hotel Maritim)

# Program of BMBF-Workshop „ErUM-Data“



UNI  
FREIBURG

Freitag, 05.10.2018

- 9:00 - 10:30    Arbeitsgruppen: Kurzvorträge, Diskussion, Gruppenarbeit  
                  PGR            AG 1: Förderierte Infrastrukturen  
                  Saal 2            AG 2: Big Data Analytics  
                  Saal 1            AG 3: Forschungsdatenmanagement
- 10:30 - 11:00    Pause (Foyer vor Saal 2)
- 11:00 - 13:00    Fortsetzung Arbeitsgruppen: Diskussion, Ergebniszusammenfassung  
                  PGR            AG 1: Förderierte Infrastrukturen  
                  Saal 2            AG 2: Big Data Analytics  
                  Saal 1            AG 3: Forschungsdatenmanagement
- 13:00 - 14:00    Mittagsimbiss (Foyer vor Saal 2)
- 14:00 - 15:30    Vorstellung der Arbeitsergebnisse (jeweils inkl. 15 Min. Diskussion)  
                  14:00            AG 1: Förderierte Infrastrukturen, *Leiter der AG*  
                  14:30            AG 2: Big Data Analytics, *Leiter der AG*  
                  15:00            AG 3: Forschungsdatenmanagement, *Leiter der AG*
- 15:30- 16:00    Pause (Foyer vor Saal 2)
- 16:00 - 16:45    Abschluss  
                  16:00            Zusammenfassung der Ergebnisse (inkl. 15 Min. Diskussion), *Sprecher der AGs*  
                  16:30            Schlusswort, *MinDirig Dr. V. Dietz, BMBF*
- 16:45            Ende der Veranstaltung

# Key Topics and Working Groups



Title of the Workshop:

„Herausforderungen und Chancen der Digitalisierung in der Grundlagenforschung“  
(Challenges and Opportunities of Digitalisation in Basic Research)

## Föderierte Infrastrukturen

Dr. Andreas Haungs	KAT	Karlsruher Institut für Technologie (KIT)
Prof. Dr. Günter Quast	KET	Karlsruher Institut für Technologie (KIT)
Dr. Marcel Kunze	KHuK	Ruprecht-Karls-Universität Heidelberg

## Big Data Analytics

Prof. Dr. Martin Erdmann	KAT	Rheinisch-Westfälische Technische Hochschule Aachen
Prof. Dr. Christian Gutt	KFS	Universität Siegen
Prof. Dr. Thomas Kuhr	KET	Ludwig-Maximilians-Universität München

## Forschungsdatenmanagement

Dr. Klaudia Hradil	KFN	Technische Universität Wien
Prof. Dr. Anke-Susanne Müller	KfB	Karlsruher Institut für Technologie (KIT)
Prof. Dr. Matthias Steinmetz	RDS	Leibniz-Institut für Astrophysik Potsdam (AIP)

This team

- organizes program of parallel sessions and coordinates and stimulates the discussion
- will be part of an editorial team providing a paper with recommendations, which shall be approved by the committees (KAT, KET, KfB, KFS, KFN, KHUK, RDS)

# Questions to the Community

## a) Federated Infrastructures



### Föderierte Infrastrukturen: Fragen zu Stand und Bedarf der Communities

Welche Infrastrukturen nutzt ihre Community zur Zeit?

- Computing Modell: Föderiert oder lokal?
- Welcher Art sind die Anforderungen, z.B. High Performance (HPC), High Throughput (HTC), Archivierung, Zugriffsmuster auf Daten, Besonderheiten ...
- Verwendung von Spezialarchitekturen: Z.B. GPUs, FPGAs, ...?

Welchen zukünftigen Bedarf sehen Sie im Vergleich zu heute?

- Rechenleistung, Datenvolumen und Netzwerkbandbreite
- Änderung der Anforderungen (in welchem Zeitraum)
- Echtzeitanwendungen (Online) vs. Stapelverarbeitung (Offline)?
- Zugriffsraten, Userdefinitionen, Zugriffsrechte

Erfahrungen mit der Nutzung von Cloud Technologien?

- Verwendung von "Science Clouds"
- Anwendungen, die in Cloud-Umgebungen laufen bzw. laufen könnten?  
Kann Container-Technologie zum Einsatz kommen (Docker, Singularity, ...) ?
- Welchen Bedarf an Unterstützung sehen Sie (Forschung und Entwicklung, Betriebs- und Anwenderschulungen, langfristiger Support, ...)?
- Notwendige Weiterentwicklung der Software-Infrastruktur:  
Community-spezifische Anpassung und Weiterentwicklung von OpenSource Komponenten, Frameworks etc.  
Optimierung von Algorithmen, Methoden, Analysewerkzeugen etc. bzgl. Multicore und Spezialarchitekturen

# Questions to the Community

## b) Big Data Analytics



### Big Data Analytics: Fragen zu Angleichung des Sprachverständnisses & Bedarfe der Communities

#### Fragestellungen und Erwartungspotential für Big Data Analytics

- Welche Fragestellungen sollen gelöst werden?
- Welche Erwartungen sind mit Big Data Analytics verbunden?
- Wie viel Erfahrung gibt es bereits?
- Bewertung vergangener Erfahrungen?
- Welche Herausforderungen sind zu lösen?

#### Datenlevel für Big Data Analytics

- Online „am Mess-Sensor“?
- Offline aus Dateien?
- Auf Einzelereignissen (z.B. Kollisionsereignisse, Röntgenbildern, ...)?
- Auf Gesamtheit von Datensätzen (z.B. sky maps, überlagerte Bilder, ...)?
- Welche Datenstrukturen und –formate werden verwendet?
- Wer speichert wo die Daten?

# Questions to the Community

## b) Big Data Analytics (2)



### Datenvolumen für „Big Data“ Analysen

- Typische Datenvolumina pro Analyse, durchgeführte Analysen pro Jahr
- Bytes (oder Anzahl Variable) pro Datensatzelement

### Analysemethoden für Big Data Analytics

- Anpassungen mit sehr großen Parameteranzahlen (z.B. Fits mit >10,000 Parametern)?
- Methoden mit großen Anforderungen von Trainingsdaten (deep networks)?
- Methoden mit numerischen Inference-Verfahren (z.B. Bayes: information field theory)?

### Existierende und erwartete Kooperationen Big Data Analytics

- Innerhalb des eigenen Community-Bereichs (z.B. nur Teilchen)
- Benachbarte Communities (Teilchen-Astroteilchen, Synchrotron-Neutronen, ...)
- Internationale Partner
- Mathematik, Informatik
- Wirtschaft

# Questions to the Community

## c) Research Data Management



### Forschungsdatenmanagement: Kernfragen zur Beantwortung durch die Komitees

#### 1. Metadaten:

Wenn Sie von den FAIR-Prinzipien ausgehen:

- können Sie die Findbarkeit in ihrer Community auf Basis der DOI/DataCite-Anforderungen organisieren
- sind die in Ihrer Community verwendeten Schemata genügend ausgearbeitet, um Interoperabilität zu sichern

#### 2. Einheitliches FDM:

- sind in Ihrer Disziplin allgemein anerkannte Workflows/Verfahrensschritte für das Forschungsdatenmanagement etabliert
- würden Sie ihre eigenen Forschungsdaten einem interdisziplinären oder einem disziplinären Forschungsdatenzentrum geben zur Archivierung und Publikation?

#### 3. Datenreduktion:

- gibt es in Ihrer Disziplin übliche Verfahren zur Reduktion von Rohdaten bzw. werden abgeleitete Daten archiviert?

# Cordial Thanks to



UNI  
FREIBURG

Local organizers: Torsten Harenberg, Marisa Sandhoff, Christian Zeitnitz

Program committee: Markus Elsing, Thomas Kuhr, Günter Quast, Kilian Schwarz



and of course to  
all participant

Looking forward to a  
lively, fruitful workshop  
with hopefully a  
consensus at the end  
and that we do not share  
the fate of Tuffi