

# Ein paar Informationen und Diskussion des ErUM-Data-Strategiepapiers

Albert-Ludwigs-Universität Freiburg

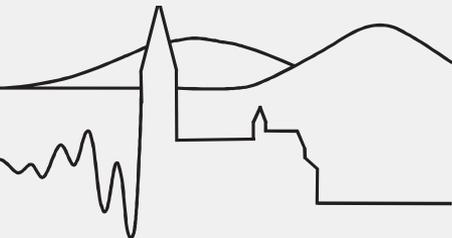


**UNI  
FREIBURG**

Markus Schumacher  
KET-C&S-Panel-Treffen  
5.4.2019

Physikalisches Institut

Albert-Ludwigs-  
Universität Freiburg





# Übersicht

- Änderungen in Zusammensetzung des KET-C&S-Panel
- Informationen zu NFDI und „Extended Abstract“ PAHN-PaN
- Zeitplan des ErUM-Data-Strategiepapiers
- Zusammenstellung des Status-Quo und Bedarf in der KET-Community
- Diskussion des ErUM-Data-Strategiepapiers in der Version vom 1. April
- Weitere Diskussionspunkte (auch für weiteres Treffen)



# Zusammensetzung des KET Computing&Software-Panel

Entsender	bis 2/2019	ab 3/2019
KET	M. Elsing	
KET	<u>M. Schumacher</u>	V. Büscher
ALICE	T. Kolleger	
ATLAS	A. Quadt	<u>M. Schumacher</u>
BELLE	T. Kuhr	
CMS	<u>G. Quast</u>	
LHCB	J. Marks	
LC	F. Gäde	
DESY	V. Gülzow	
KIT	A. Heiss	
MPG	S. Kluth	
GridKa TAB	K. Schwarz	J. Marks
GridKa TAB	T. Kress	T. Harenberg
KHUK-Data		K. Schwarz

- Theorie bisher nicht vertreten.  
KET ist dabei 1 oder 2 Vertreter bis Sommer zu bestimmen
- Vielen Dank an  
Arnulf Quadt und Thomas Kress für Ihre Mitarbeit.
- Herzliches Willkommen an  
Thorsten Harenberg und Volker Büscher
- Kilian ist Vertreter von KHuK-Data im KET-C&S-Panel
- Jörg Marks ist Vertreter von KET-C&S in KHuK-Data
- Eventuell Vorsitz neu bestimmen/wählen?



Bund-Länder-Vereinbarung zu Aufbau und Betrieb einer NFDI vom 28. 11. 2018

## Ziele der Förderung

- Etablierung und Fortentwicklung eines übergreifenden Forschungsdaten-management u. Steigerung der Effizienz des gesamten Wissenschaftssystems
- a) Aufbau einer koordinierten, vernetzten Informationsinfrastruktur zur Entwicklung eines nachhaltigen interoperablen Forschungsdatenmanagement
- b) Etablierung von in den wiss. Disziplinen akzeptierten Prozessen und Verfahren zur standardisierten Umgang mit Forschungsdaten
- c) Schaffung eines verlässlichen und nachhaltigen Dienste-Angebots, ....
- d) Etablierung disziplinübergreifender Metadatenstandards ...
- e) Anbindung der deutschen FDI an eur. und int. Plattformen
- f) Optimierung der Nutzbarkeit, ..., der Infrastrukturen , ...
- g) Schaffung einer gemeinsamen Basis für Datenschutz sowie Souveränität, Integrität, Sicherheit und Qualität von Daten

Gegenstand der Förderung: Konsortien (ca. 30 in 3 Antragsrunden) und Direktorium

Nur Personal. Kein Invest in Hardware Förderzeitraum: 2019 bis 2028

Fördervolumen 90 MEuro/Jahr (85MEuro für Konsortien, 5 MEuro für Direktorium)

Erwartetes Volumen pro Konsortium 2-5MEuro/Jahr Förderbeginn Sommer 2020

Mehr Infos unter <http://www.dfg.de/foerderung/programme/nfdi/>



# Zeitplan

29. März 2019 Frist für Einreichung von Extended Abstracts für NFDI-Konferenz
- 13.-14. Mai 2019 NFDI-Konferenz
- Juni 2019 Veröffentlichung der Ausschreibung
- Juni/Juli 2019 Frist für Einreichung von verbindlichen Absichtserklärungen
- Oktober 2019 Frist für Einreichung von Anträgen
- November 2019 – Januar 2020 Begutachtung der Anträge der ersten Runde
- März - April 2020 Kommunikation der Begutachtungsergebnisse an die Konsortien  
mit Möglichkeit zur Stellungnahme
- April 2020 Sitzung des Expertengremiums / Empfehlungen an die GWK
- Juni 2020 Erste Förderentscheidung der GWK  
(GWK - Gemeinsame Wissenschaftskonferenz)

## **NFDI-Konferenzen:**

Erster Schritt auf dem Weg zur Etablierung eines verteilten Netzwerks von fachlich oder methodisch organisierten Konsortien zu Beginn jeder Auswahlrunde .

Teilnahme ist ein Angebot, aber weder Antragsvoraussetzung noch begutachtungsrelevant.

## **Ausschreibung und Antragstellung**

Im Anschluss an Konferenz wird Ausschreibung für die aktuelle Aufnahme phase veröffentlicht. Konsortien bewerben sich mit einem schriftlichen Antrag um die Aufnahme in die NFDI.

## **Begutachtungspanels**

Der Begutachtungsprozess sieht einen intensiven Austausch zwischen Gutachtenden und Antragstellenden vor. (Anhörung der Antragstellenden, die Diskussion mit Gutachtenden)

## **Stellungnahmen der Konsortien**

Als Rückkoppelung u. Qualitätsverbesserung wird den Konsortien das Begutachtungsergebnis mitgeteilt werden, mit der Möglichkeit eine Stellungnahme dazu abzugeben.

## **Bewertung durch das NFDI-EG**

Das Expertengremium wird auf Basis der Begutachtung sowie der ggf. dazu eingereichten Stellungnahmen der Antragstellenden alle Anträge vergleichend diskutieren und bewerten.

# NFDI-Konferenz 13./14. Mai in Bonn



**Vorraussetzung: Einreichung eines „extended abstract“ bis 29.3.2019**

## **Zielsetzung**

Die NFDI-Konferenz verfolgt drei Zielsetzungen:

- a) DFG wird die Vertreter\*innen und Vertreter von geplanten Verbänden über das Konzept und übergreifenden Ziele der NFDI sowie über die Eckpunkte der Antragstellung informieren.
- b) Geplanten Verbände sollen sich untereinander und mit internationalen Expertinnen und Experten austauschen.
- c) Konferenz unterstützt die Selbstorganisation und Vernetzung der Konsortien zu konsortienübergreifenden Diensten, zu fachnahen oder datenähnlichen Angeboten, ...  
→ dort und im Nachgang wird Frage der NFDI4Physik diskutiert

## **Vorläufiges Programm**

Begrüßung / Allgemeine Informationen

Vorstellung der geplanten Konsortien in parallelen Sitzungen

Podium und Diskussion im Plenum      Marktplatz / Posterdiskussion

Eindrücke vom ersten Tag und Einführung in die Parallelgruppen

Austausch u. Diskussion zu Selbstorganisation, Vernetzung, zu Querschnittsthemen

Zusammenführen der Ergebnisse und Vorbereitung von Ergebnispräsentationen

Vorstellung der Ergebnisse der Parallelgruppen im Plenum, Resümee und Abschluss

### PANH-PaN

Particle, Astroparticle, Hadron and Nuclear Physics accelerates the NFDI

32 Partner/Institute aus KET, KHUK + (KAT ohne Astronomieanteil) + CERN

Leitende Institute: DESY und GSI

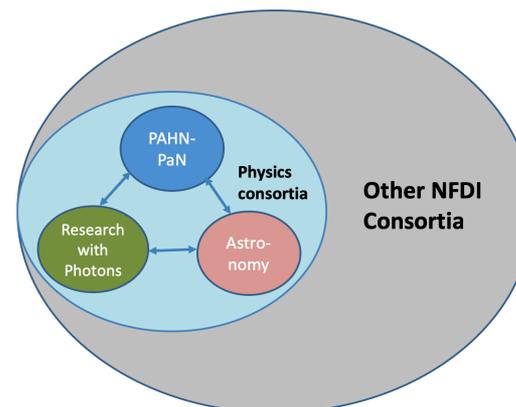
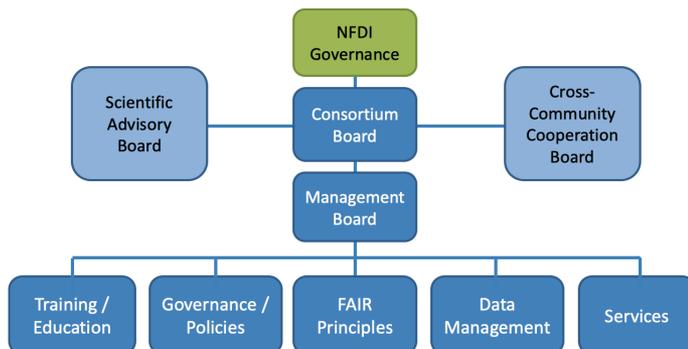
Editoren des „Extended Abstract“ K. Schwarz (GSI) und T. Schörner-Sadenius (DESY)

Abstract in verschiedenen Gremien diskutiert Fristgerecht am 29.3. eingereicht

Abstract wurde angenommen und Einladung zur NFDI-Konferenz ist gestern erfolgt

3 Teilnehmer pro Konsortium in NFDI-Konferenz am 13./14. Mai

T. Schörner-Sadenius (DESY), Kilian Schwarz (KHUK und GSI), T. Kuhr (KET)



# Zeitlinie des Workshopabschlusspapiers

- Workshop: BMBF Bonn 4./5. Oktober  
Herausforderungen u. Chancen der Digitalisierung in der Grundlagenforschung“  
(beteiligte Komitees KAT, KET, KfB, KFN, KFS, KFSi KHuK, RDS)
- Abschlusspapier mit Editorenteam aus 9 Leuten  
aus KET: T. Kuhr , G. Quast **DANKE!**                      Vorsitzender: M. Erdmann
- Erste Version 4. November- 3 Tage für Kommentierung durch Volker Gülzow, M.S.
- 9. November an BMBF/Referat 711 für Kommentierung
- 4. Februar: neue Version erstmals mit Zahlentabellen zu Invest, FTE an Komitees
- Seitdem lebhaft Diskussion und Kommentierung in Komitees
- KET-Kommentare am 26.2. an Editoren übermittelt / Bedarfszahlen am 19.3.
- Nächste Version am 1.4. an Komitees → finale Kommentare bis 8.4.
- **Heute (5.4) abschließenden Diskussion im KET-C&S-Panel**
- Montag (8.5) Übermittlung der eventuellen Kommentare durch KET an Editorenteam
- Mittwoch (10.5) gemeinsame Diskussion der Komiteevorsitzenden und Editoren
- ???
- Übergabe und Präsentation während PrismaDialog-Treffens am 2. Mai im BMBF?

# Kommentare zum Papier: Einleitung



- KET begrüßt und unterstützt die Initiative und sieht die große Chance
  - Sicht einiger/vieler: KET hat den Prozess konstruktiv begleitet
  - KET-Kommentare beabsichtigen die Interessen der KET-Community auszudrücken
  
  - Am 8.4. wird der KET-Vorsitzende finale Kommentare übermitteln
    - Ziel: wenige wichtige, die auch im Treffen am 10.4. diskutiert werden können
    - Editorische Kommentare als Anhang oder direkt an KET-Vertreter im Editorenteam
  
  - Jetzt: Diskussion der inhaltbezogenen Kommentare und  
Zusammenstellung von zentralen Kommentaren und Empfehlung für KET
- 3 Sätze von Kommentaren sind bisher eingegangen (Markus E., Stefan K, Markus S.)  
Fast komplette E-Mails in Dokument auf Webseite abgelegt

# Kommentare zum Papier: Markus E.

- 1.196: term “Smart Data” is not introduced and the context here therefore stays vague. Either drop this or introduce this term with a definition of what this should mean in the context here. (M.S.: Big data also no defined )
- 1.232: MPI as well provides computing infrastructure, should we add this ? (M.S.: should be added)
- 1.608: I do have some issues with calling "but also the analysis tools and used libraries" inside "meta data". I would have the tendency to drop this half-sentence. Analysis tools and libraries are provided on most systems by other means than (meta) data. Systems like CVMFS and alike are used for this, and so are container solutions. I guess this is a questions of where to put the boundary... (M.S.: for long term archiving also tools and libraires ar eneeded?)
- 1.616: the first bullet gives the impression that FAIR “only” means a DOI. A DOI is part of making the data FAIR, but that is obviously not sufficient. I would use the DOI as an example of what is required and write it like this. (M.S: what exactly is the suggestion?)
- 1.619: I would reorder the bullets, putting the third one as second in the list. (M.S.: why?)
- 1.808ff: Here we only mention HPC as a system providing computing resources. It provides very powerful resources of a certain type, but we should mention this at the same level as HTC and classical GRID/cloud type of resources, not to give the impression we would favour a concentration on HPC investments.  
(M:S.: HPC is only mentioned in the context of Gauss-Alliance. This is correct or not? for university clusters no speficiation w.r.t. HPC and/or HTC is made?)

# Kommentare zum Papier: Markus E. (2)

But 5.2 (l.899 ff) does not mention that workflow and data management services are developed in an EU and international context, we had e.g. mentioned EU ESCAPE project (DOMA in particular) in the text above section 5. We can/should benefit from the EU/international collaboration across domains (a la ESCAPE) and aim at even take leading roles where appropriate. [MS:?](#)

Even more so for section 5.3 (part about challenges and responsibilities), which is maybe not the strongest in terms of justification. I would prefer to talk as well here about the development, deployment and operation of services for data management and workflow management. Like in section 5.2, the EU and international context for such developments are missing, ESACPE/DOMA and e.g. Rucio as a system used by more and more of our communities. [MS:?](#)

l.966: I really don't like "committees to design the data" - this never ever worked for any cross community projects I am aware of. Data structures need to be transparent and appropriate for the domain problem. Experiments will deal with this, not an external committee that is bound to lack experiment specific expertise. For cross-experiment data sharing one needs to make data structures public and usable, not force experiments to use structures that don't match their needs. [MS:?](#)

l.1042 and l.1159: I would also drop "including physics, chemistry, biology, medicine, earth science" as others commented already, this is not precise/misleading and not needed here. → see comments by M.S.

l.1165: "brain drain" is not what I would write here, it is a goal that we educate data scientists for industry. But the balance is not there, we risk to lose the experts faster than we can educate new ones. I would formulate a bit more careful. [M.S.: how exactly?](#)

l.1320: I thought we do not include Fair Tier-0 and SKA resources in the numbers here, but then it should be said? [MS: see line 856—859, should it really be repeated here?](#)

# Kommentare zum Papier: Stefan K.



Hier nur konkrete Kommenatre aus 2. E-Mail

Anmerkunge aus 1. E-Mail gut udn wichtig aber zu allgemeine oder zu fundamentale Kommentare für derzeitigen Status, aber für weitere Diskussionen nach Papier archivieren,

Generell, ueberall wo "federated infratstructures" steht, nur"infrastructures" schreiben. Wir haben keine Definition des Begriffs"federated infrastructure" und so besteht die Gefahr das jede community das anders versteht. Und wir sollten nicht die Conclusion (das wir federated infrastrucuters (im HEP Sinn) brauchen) an den Anfang stellen, bevor wir die Fakten gesammelt und ausgewertet haben. ....  
→ Siehe Antwort von Günter Q. Begriff war eine Arbeitsgruppe beim Workshop

l 52 ff: New funding schemes are necessary to enable this development, thus allowing to exploit the full potential of the data provided by the next generation facilities. **Such changes must be well coordinated with other programs concentrating on establishing novel infrastructures such as the National Research Data Infrastructure (NFDI) and the European Open Science Cloud (EOSC). Our work on infrastructures for future ErUM research in a evolving world of Big Data and Big Data Analytics can thus be seen and used as input by the ErUM community for the development of the NFDI. (S.K. suggests to add)**

l 76 ff ~~Federated~~-Infrastructures: provisioning of ~~federated~~-infrastructures "beyond Moore" to enable the repeated processing, training of Big Data Analytics tools, and archiving of data volumes in the exabyte range including the required high-throughput network infrastructure. **(S.K. suggest to delete „federated“)**

Siehe obigen Kommentar

# Kommentare zum Papier: Stefan K. (2)

1 370 ff Ein extra bullet: **consideration of the special demands of training of Big Data Analytics tools with the expected data volumes in the Exabyte range (S.K. add)**

1 403 ff bzw 1 405 ff nur ein bullet (die beiden vorhandenen koennen auch widerspruechlich gelesen werden)  
 o **Creation of a federation of computer centres serving as a "ErUM Science Cloud" with integration into national concepts such as the "National Research Data Infrastructure" as well as EU-funded endeavors such as EOSC. (S.K. combine bullets)**

1 419 ff The infrastructure needs to meet the future challenges in terms of storage, (German share of data volumes of exabyte per year), network (high-bandwidth network with 100 gigabit per second short-term and terabit per second long-term) connecting all ErUM centres, computing power **for data processing, simulation and Big Data Analytics tools training, and services. (S.K. add)**

1 727 ff o **Federated** Infrastructures: Owing to the rapid increase in data volumes and the rise of Big Data Analytics tools in the coming years, the demand for compute power, storage space and networks will increase dramatically. We see an immediate need for action **in the well-planned and significant expansion** of the infrastructure.(S.K. **delete** and **add**)

1 786 am Ende des Absatzes dazu: **In addition, the training of Big Data Analytics tools on massive amounts of data, which is a rapidly evolving field of research, will very likely need very large and specialised compute resources. (S.K add)**

# Kommentare zum Papier: Markus S.



I) Aus meiner Sicht unbedingt notwendig sind Änderungen bei:

a) In der Tabelle auf dem Deckblatt steht „on behalf of Committee“

Es war der explizite Wunsch von KET, dass dies so nicht erscheint.

Der Ausdruck spiegelt nach Meinung mehrerer Leute auch nicht den Prozessverlauf wieder.

→ Wird im Treffen der Komiteevorsitzenden mit Editorenteam am 10.4. diskutiert.

b) Abschnitt 5 Zeile 706 bis 708: „extensive survey ... in-depth analysis ...“

Das hat meiner Meinung nicht stattgefunden und sollte so definitiv auch nicht erscheinen.

Ich schlage vor, dass hier ein andere Formulierung gewählt wird.

II) Es wäre wünschenswert auch folgende Passagen noch klarer zu gestalten:

a) bei 5.4. steht explizit „long-term position“ aber z.B. bei 5.2 nicht. Was genau ist mit „long-term position“ gemeint? Warum wird hier ein Unterschied in der Bezeichnung in 5.4. und an anderer Stelle z.B. 5.2 gemacht? Ist die Unterscheidung beabsichtigt und falls ja, wie ist der Unterschied begründet?

b) an mehreren Stellen steht "ErUM-Communities including physics, medicine, biology, chemistry" (z.B. Zeilen 1041 und 1159,). Das ist meiner Meinung nach nicht korrekt. ErUM umfasst im wesentlichen die 8 Komitees und nicht auch die anderen Forscher\*innengemeinden. Ich fände eine andere Formulierung wünschenswert. (ähnlicher Kommentar von Markus E.)

→ Thomas Kuhr hat geantwortet und Änderung ist für MS nicht wichtig.



# Weitere Diskussionspunkte oder Denkanstöße für nächstes Treffen

- Weitere Änderungen in Zusammensetzung des KET-C&S-Panel wünschenswert?
- Austausch/Neubestimmung des Vorsitzenden (und seines Stellvertreters)?
- Regelmäßige Treffen wie KET?
- Anfertigung/Finalisierung eines KET-Papiers zu Computing&Software ?  
(Dank an Manuel G., Thomas. Ktress., Kilian S., Günter Q., ... )