

universität freiburg

Einleitung zum Planungstreffen für Verbundantrag Nf FIDIUM

Markus Schumacher
Universität Freiburg

14. April 2023



Status Ausschreibung, Empfehlungen und Zeitskalen

Ausschreibung im Programm ErUM-Pro im Bereich „Teilchen“ noch nicht veröffentlicht
Text ist seit Wochen fertig aber hängt noch in der BMBF-Bürokratie

Veröffentlichung wird in den „Nächsten Tagen“ erwartet. Im Text steht wohl etwas zu „Datenmanagement etc“.

Auch in inoffiziellen Gesprächen mit PT-DESY noch keine Aussage, ob Einreichung eines Antrages
„Nf FIDIUM“ von PT-DESY empfohlen bzw. erwünscht wird
(Alternative ist Laufzeitverlängerung (um 1 Jahr) von FIDIUM mit zusätzlichem Geld)

Falls Neuantrag wird Laufzeit des Projektes wohl 10/2024 bis 6/2027 sein, da FIDIUM bis 9/2024 läuft.

Letztes Telefonat heute Vormittag mit Wolfgang Ehrenfeld hat keine neuen Erkenntnisse gebracht.
Hoffentlich ist PT-DESY in den nächsten 10 Tagen sprechfähig.

→Wir sollten dennoch jetzt an Vorbereitung eines Verbundantrages arbeiten und
für Einreichung bei der ErUM-Pro-Ausschreibung zum 30.6 / 1.7. 2023 bereit sein.

Erinnerung: FIDIUM Partner

Föderierte Digitale Infrastrukturen für die Erforschung von Universum und Materie (FIDIUM)

Gemeinsamer Antrag von Gruppen aus den Bereichen Elementarteilchenphysik, Hadronen-
und Kernphysik und Astroteilchenphysik

- Rheinisch-Westfälische Technische Hochschule Aachen, Prof. Dr. Alexander Schmidt¹
- Rheinische Friedrich-Wilhelms-Universität Bonn, PD Dr. Philip Bechtle
- Goethe Universität Frankfurt am Main, Prof. Dr. Volker Lindenstruth
- Albert-Ludwigs-Universität Freiburg, Prof. Dr. Markus Schumacher
- Georg-August-Universität Göttingen, Prof. Dr. Arnulf Quadt
- Universität Hamburg, Prof. Dr. Johannes Haller
- Karlsruher Institut für Technologie, Prof. Dr. Günter Quast
- Johannes Gutenberg-Universität Mainz, Prof. Dr. Frank Maas
- Ludwig-Maximilians-Universität München, Prof. Dr. Thomas Kuhr
- Bergische Universität Wuppertal, Prof. Dr. Christian Zeitnitz

Assoziierte Partner sind

- CERN, Dr. Markus Elsing
- DESY, Prof. Dr. Volker Gülzow
- GridKa, Dr. Andreas Petzold
- GSI Helmholtzzentrum für Schwerionenforschung, Dr. Kilian Schwarz²

Neue Partner?

TUM München, Lukas Heinrich?

Frankfurt 2, Thorsten Kollegger

Weitere?

Koordinatoren/designierte Sprecher:

Markus Schumacher (KET-Bereich)

Thorsten Kollegger (KHuK-Bereich)

CERN und GSI haben Interesse geäußert
weiterhin assoziierte Partner zu sein
auch wenn heute kein Vortrag gehalten wird

Erinnerung: FIDIUM Themenbereiche und Arbeitspakete

3.1 Themenbereich I: Entwicklung von Werkzeugen zur Einbindung heterogener Ressourcen

1. Erschließung und effiziente Einbindung von opportunistischen Ressourcen

- Weiterentwicklung und Anpassung des Resource Managers COBa1D/TARDIS an zukünftige Gegebenheiten
- Entwicklung von dynamischer Steuerung des Job-Schedulings (z.B. Berücksichtigung von Datenlokalität, I/O-Raten)
- Automatisierte Skalierung peripherer Dienste
- „Compute Site in a Box“: Nutzbarmachung der Ressourcen an Tier-3- und Tier-2-Zentren mit minimalen zusätzlichen administrativen Ressourcen (volle Automatisierung, Skalierbarkeit)

2. Accounting und Controlling von heterogenen Ressourcen

- Werkzeuge zum Accounting der opportunistisch genutzten Ressourcen
- Tools für kontinuierliche Überwachung der Nutzungseffizienz

Die beteiligten Partner im Themenbereich, sowie die beantragten Mittel sind in Tab aufgeführt. Die Koordination des Themenbereichs wird übernommen von Dr. Manuel C und Dr. Oliver Freyermuth.

Standort	PI	FTE	Experiment	AP 1	AP 2
KIT	G. Quast / A. Streit	0.66	CMS	X	X
U Bonn	P. Bechtle	1	ATLAS/Belle II	X	
GU Frankfurt	V. Lindenstruth	0	ALICE/CBM	X	
U Freiburg	M. Schumacher	1.2	ATLAS	X	X
U Göttingen	A. Quadt	0.5	ATLAS	X	
U Wuppertal	C. Zeitnitz	0.5	ATLAS		X
Assoziiert					
GSI	K. Schwarz	-	ALICE	X	
DESY	V. Gülzow	-	verschiedene	X	
GridKa	A. Petzold	-	verschiedene	X	X

Erinnerung: FIDIUM Themenbereiche und Arbeitspakete

3.2 Themenbereich II: Data-Lakes, Distributed Data, Caching

1. Aufbau eines Echtzeit Data-Lake-Monitoring-Systems

- Erfassung der Auslastung von Data-Lake-Komponenten
- Erfassung von Datenzugriffsmustern

2. Technologien für Data-Lake-Caching

- Weiterentwicklung und Konsolidierung von Daten-Cache - Technologien
- effiziente Einbindung von dynamischen Datencaches in den Data-Lake und an CPU-Ressourcen
- Einsatz von parallelen ad-hoc Filesystemen als Caches in HPC-Systemen

3. Technologien für Data-Lake-Daten- und Workflow-Management

- Replikations- und Platzierungsmechanismen
- bedarfsgetriebene Datenmanagement-Mechanismen
- effizienter Datenzugriff und Anpassung an Workload-Management-Systeme

4. Data-Lake-Prototypen, Technologien für QoS und effiziente Anbindung

- Aufbau von Data-Lake-Prototypen
- effiziente Anbindung von Nutzern, Zentren und Datenquellen
- Quality of Service

Die beteiligten Partner im Themenbereich, sowie die beantragten Mittel sind in Tabelle 2 aufgeführt. Die Koordination des Themenbereichs wird übernommen von Dr. Kilian Schwarz und Prof. Dr. Andre Brinkmann.

Standort	PI	FTE	Experiment	AP 1	AP 2	AP 3	AP 4
KIT	G. Quast / A. Streit	0.66	CMS		X	X	
KIT	R. Engel	0.5	Auger/Einst.- Tel./IceCube	X		X	X
GU Frankfurt	V. Lindenstruth	1	ALICE/CBM		X		X
U Mainz	F. Maas / A. Brinkmann	1	PANDA		X	X	X
LMU München	G. Duckeck	1	ATLAS		X		X
U Hamburg	J. Haller	0.66	CMS		X	X	
U Göttingen	A. Quadt	0.5	ATLAS			X	
U Wuppertal	C. Zeitnitz	0.5	ATLAS	X			
Assoziiert							
GSI	K. Schwarz	-	ALICE	X	X	X	X
CERN	M. Elsing	-	ATLAS			X	X
DESY	V. Gülzow	-	verschiedene		X	X	X
GridKa	A. Petzold	-	verschiedene		X	X	X

Erinnerung: FIDIUM Themenbereiche und Arbeitspakete

3.3 Themenbereich III: Anpassung, Test und Optimierung auf Produktions- und Analyse-Umgebungen

1. Integration, Tests, Optimierung und Deployment der entwickelten Dienste

- Integration der verschiedenen Komponenten: Workflowmanagement, Caching, Accounting, Ressourcenmanagement und Überwachungssysteme
- Funktionale Tests auf ausgewählten Zentren
- Integration in die Produktionsumgebung der beteiligten Experimente
- Am Ende, nach erfolgreichen Tests, "Deployment" der Gesamtlösung an den verfügbaren WLCG-Tier-Zentren, HPC-Zentren und Cloud Anbietern.

2. Spezifische Anpassung der Dienste an komplexe Workflows und Nutzung spezieller Technologien für die Analyse wissenschaftlicher Daten

- Optimierung für spezielle Workflows mit hoher IO Last, Speicherbedarf, GPU Nutzung, u.a.
- Optimierung für schnelle parallele Analyse großer Datenmengen mit modernen vektor-basierten Analysealgorithmen

3. Support

- Einrichtung eines standortübergreifenden Support-Teams, das die Zentren bei Installation und Betrieb unterstützt

Die beteiligten Partner im Themenbereich, sowie die beantragten Mittel sind in Tabelle 3 aufgeführt. Die Koordination des Themenbereichs wird übernommen von Prof. Christian Zeitnitz und Dr. Günter Duckeck.

Standort	PI	FTE	Experiment	AP 1	AP 2	AP 3
RWTH Aachen	A. Schmidt / M. Erdmann	2	CMS / Einstein Teleskop	X	X	
KIT	A. Stahl					
KIT	G. Quast / A. Streit	0.66	CMS	X	X	X
KIT	R. Engel	0.5	Auger/IceCube/Einstein Teleskop	X		X
Uni Mainz	F. Maas / A. Brinkmann	0.66	PANDA	X		
Uni Wuppertal	C. Zeitnitz	0.66	ATLAS	X		
GU Frankfurt	V. Lindenstruth	1	ALICE/CBM	X	X	
LMU München	T. Kuhr / G. Duckeck	1	Belle II / ATLAS	X	X	
U Freiburg	M. Schumacher	0.8	ATLAS	X		X
U Hamburg	J. Haller	0.66	CMS	X	X	
Uni Göttingen	A. Quadt	0.66	ATLAS	X		
Assoziiert						
GSI	K. Schwarz	-	ALICE	X		X
DESY	V. Gülzow	-	verschiedene	X	X	X
GridKa	A. Petzold	-	verschiedene	X	X	X

Für Neuantrag sind modifizierte und neue Themenbereiche willkommen

Gesamtanzahl sollte nach Meinung der Koordinatoren aber maximal vier sein.

Erste Ideen für NF FIDIUM im E-Mail-Verteiler

Alexander S. bzw Kilian S.

- Erlernen des Betriebs der zentralen Daten-Cloud durch HGF-Zentren
- Erlernen des Betriebs der IT-Ressourcen durch die HPC-Zentren
- Verwendung von Beschleunigern (GPU, FPGA ...)
- Weiterentwicklung der Software, die in FIDIUM entstand

Günter Q. und Manuel G.

- Themenliste unten (→hier oben) ist aus unserer Sicht am KIT schon recht vollständig - über Wording sollten wir allerdings noch sprechen ("Erlernen" -> "Support durch HEP-ExpertInnen" usw.).
- Wir sollten auch andere als LHC-Experimente (Belle II, evtl. IceCube und Auger explizit erwähnen, auch aus diesen Communities gibt es erste Kundschaft für CoBALD/Tardis).
- **Wir sollten uns auch über den Personalbedarf verständigen. Große HPC-Ressourcen benötigen in der Übergangsphase sicher je 1/2 FTE, angesiedelt an einem HEP-Institut.**
- **Einige Ressourcen sollen ja von der HEP zentral oder an Tier2 betrieben werden. Die betreffenden Institutionen brauchen auch dafür Personal (momentan ist das für CMS bei GridKa der Fall, spezielle Set-Ups für z.B. ATLAS brauchen aber weiteren personellen Support).**
- Als weiteren Punkt sollten wir noch "Anbindung von universitären Analyseclustern" als Punkt hinzunehmen; einerseits gibt es dort opportunistische HPC-Ressourcen, andererseits müssen sie mit Daten versorgt werden und benötigen evtl. lokale Caches, die man aber regional sichtbar machen und verwenden sollte.

MS Anmerkung: Aufgaben in „Rot“ gehören m. M. nach in Verbund zu Betrieb und experimentspezifischen Entwicklungen

Erste Ideen für NF FIDIUM im E-Mail-Verteiler (2)

Andreas Redelbach

- kürzlich habe ich bei einem LHC-Experiment gesehen, dass Strategien zum Green Computing konkreter werden und mehr in den Fokus rücken. Zwei Veröffentlichungen sind mir in dem Bereich aufgefallen, die sicher auch für FIDIUM 2 interessant sein könnten: <https://arxiv.org/abs/2203.12389v2> <https://arxiv.org/abs/2007.07610>
- Optimierungen in Richtung Green Computing
- Machine Learning im Ressourcenmanagement

MS Anmerkung: Nachhaltigkeitsaspekte wurden von Herrn Lilienthal beim PRISMA-Strategiegespräch „Teilchen“ sehr betont und solche Aspekte im Neuantrag sind für Erfolg sicherlich nicht unwichtig.

weitere Ideen aus bilateralen Gesprächen:

- Entwicklungen zum Aufbau von dynamischen interaktiven Analyseplattformen
- Anpassung, Wartung von entwickelten Tools und Nutzerbetreuung für viele Leute wichtig
- Ausrollen der entwickelten Konzepte und Technologien in andere Wissenschafts-Communities befördern
- Verbessertes Energiemonitoring der Jobs / weitere Nachhaltigkeitsaspekte

Ziel für heute

Unabdingbar: - Überblick über Pläne / Ideen in den einzelnen Gruppen bekommen
- Kooperationen und Interessenüberlapp identifizieren

Falls möglich: - vorläufige Themenbereiche identifizieren
- Namen von möglichen Koordinator*innen ad-hoc oder im Nachgang finden
- erste vorläufige und unvollständige Liste von möglichen Arbeitspaketen erstellen

Hoffentlich: - Zeitplan für weitere Schritte entwickeln
- Weitere Treffen der Themenbereichskoordinatoren und Interessierten anbahnen

Für Neuantrag sind modifizierte und neue Themenbereiche willkommen
Gesamtanzahl sollte nach Meinung der Koordinatoren aber maximal vier sein.