

Ziele und Verbindung zu anderen Projekten:

- Gewinn für einige der im Moment mehr heterogenen arbeitenden Forschungsrichtungen durch größere gemeinsame Infrastrukturen zum Management von Wissenschaftsdaten
- (wo möglich) gemeinsame Standards etablieren, die Interoperabilität ermöglichen
- zukünftige Entwicklungen sollen von den Erfahrungen aus existierenden Systemen zum Management sehr großer Datenmengen innerhalb von ErUM profitieren, ebenso sollen auch Kollaborationen mit Computerwissenschaften und anderen verwandten Forschungsrichtungen gesucht werden,
- Systeme zum leichten und automatisierten Zugriff auf Wissenschaftsdaten zur Datenanalyse
- “Data Life Cycle Managements” und Metadaten sind von zentraler Bedeutung
- Systeme zur Verwaltung von Exa-Scale Datenmengen und von heterogenen Daten, inklusive des Aspekts der Archivierung und Nutzbarmachung der Daten für zukünftige Forschung, sowie des Open Access
- die Schulung und das Training von Spezialisten und Wissenschaftlern im Bereich des Managements von Forschungsdaten von entscheidender Bedeutung ist.

FAIR – NFDI

TB – RDM

TB - FI

Ziele und Verbindung zu anderen Projekten:

- Gewinn für einige der im Moment mehr heterogenen arbeitenden Forschungsrichtungen durch **größere gemeinsame Infrastrukturen zum Management von Wissenschaftsdaten**
- (wo möglich) **gemeinsame Standards etablieren, die Interoperabilität ermöglichen**
- **zukünftige Entwicklungen sollen von den Erfahrungen aus existieren Systemen zum Management sehr großer Datenmengen innerhalb von ErUM profitieren**, ebenso sollen auch Kollaborationen mit Computerwissenschaften und anderen verwandten Forschungsrichtungen gesucht werden,
- Systeme zum leichten und automatisierten Zugriff auf Wissenschaftsdaten zur Datenanalyse
- **“Data Life Cycle Managements” und Metadaten sind von von zentraler Bedeutung**
- **Systeme zur Verwaltung von Exa-Scale Datenmengen und von heterogenen Daten**, inklusive des Aspekts der **Archivierung und Nutzarmachung der Daten für zukünftige Forschung, sowie des Open Access**
- die Schulung und das Training von Spezialisten und Wissenschaftlern im Bereich des Managements von Forschungsdaten von entscheidender Bedeutung ist.

Mit Komitees arbeiten
oder mit wiss. Namen?

Derzeitiger Stand in den Communities

- KET, KAT, KHuK – Exa-Scale Daten mit hochgradig standardisierter, automatisierter Verwaltung / Bearbeitung
- KfB – ebenso mit anderen Raten aber in-situ
- RDS – große Bandbreite an Arbeitsweisen, wie oben bis individuell
- KfS, KFN, KfSI – bisher Erzeugung der Daten an Zentren und Verarbeitung in den Forschergruppen; im Wandel durch höhere Datenmengen und –raten, aber noch keine etablierten Strukturen und Verantwortlichkeiten

→ Breite an Förderbedarfen

- High-end Entwicklungen auf der einen Seite, um Fortschritt zu sichern
- Entwicklung von angepasstem Datenmanagement an die Bedingungen von Nutzerbetrieb in Zentren unter Übertragung vorhandener Erfahrungen

Vorschlag Themenbereiche:

1. Entwicklung von (modularen) Systemen zum Forschungsdatenmanagement. Hierbei sollte ein Fokus auf die Verwendung und Anpassung von bereits existierenden Systemen oder die Integration solcher Systeme in wissenschaftsspezifische Lösungen gelegt werden. Mögliche Förderschwerpunkte umfassen beispielsweise folgenden Bereiche:

- Einheitliche Standardlösungen für das Management von heterogenen Wissenschaftsdaten, die insbesondere der äußerst ausgeprägten Diversität innerhalb und zwischen den ErUM-Communities Rechnung tragen.
- Interfaces und Programme, auch User Interfaces (einschließlich. Web- oder Kommandozeilen-basiert) zum leichten Zugriff auf und zum Verwalten von großen Datenmengen und Metadaten. Dazu gehören insbesondere Systeme, die die Umsetzung der FAIR-Prinzipien ermöglichen (z.B. Bereitstellung von "ZENODO-ready"-Daten) und einen einheitlichen Zugriff und die Auswertung verschiedener Experimente erlauben.
- Interfaces und Programme zur Katalogisierung der diversen, in großen Mengen anfallenden Daten und zu deren Auffindbarkeit (eigene und fremde Daten).
- (Community-übergreifende) Plattform zum Hinzufügen und Abrufen von Trainings- und Validierungsdatensätzen für maschinelles Lernen (analog / komplementär zu Referenzdatenbanken) .

Vorschlag Themenbereiche:

2. Entwicklung von Systemen zur Verwaltung und zum Zugriff auf Metadaten. Mögliche Förderschwerpunkte umfassen beispielsweise folgenden Bereiche:

- Modulare, homogenisierte und community-übergreifende Systeme zum Management und zur Dokumentation der Metadaten sowie Daten-Provenienz. Metadaten inkludieren Parameter für die Datenreduktion und Datenauswertung. Insbesondere sollte sichergestellt werden, dass die einzelnen Schritte von Rohdaten zur Abbildung in der Publikation nachvollziehbar sind. Idealerweise sollten die Metadaten auch (Vor)-Experimente von Laboreinrichtungen beinhalten.
- Workflows zur flexiblen Gestaltung der Metadatenstruktur für spezifische Instrumente und/oder Experimente.

Vorschlag Themenbereiche:

3. Experimenten und deren Steuerung. Mögliche Förderschwerpunkte umfassen beispielsweise folgenden Bereiche:

- Workflow Systeme zur Reduktion von (Roh-)Daten durch Anwendung von speziellen Algorithmen. Hierzu gehören z.B. Klassifizierung, Auswahl, und Ausrangieren von Daten sowie die entsprechenden Metadaten.
- Systeme zur effizienten Datenkomprimierung (verlustfrei oder nicht verlustfrei).
- Workflow Systeme zur Verarbeitung und Auswertung von Daten ähnlicher Experimente von verschiedenen Großforschungseinrichtungen.
- Monitoring und Validierung der Daten selbst, als auch Monitoring von Daten-Workflows und der Konsistenz (Datenverlust und Replika Management) von sehr großen Datenmengen.
- Workflow Systeme für die effiziente und automatische Visualisierung, Organisation und Extraktion relevanter Informationen aus großen Datensätzen, um Experimente zu steuern und fortzuführen.

Vorschlag Themenbereiche:

4. Entwicklung von Systemen für effizientes Life Cycle Management. Mögliche Förderschwerpunkte umfassen beispielsweise folgenden Bereiche:

- Automatisierung von Replika-Management und Strategien zur Optimierung der Verfügbarkeit relevanter und interessanter Daten.
- Strategien und Systeme zum (semi-)automatischen Löschen von sehr großen Forschungsdaten.
- Strategien und Systeme zur optimalen Ausnutzung verfügbarer Computing-Ressourcen (insbesondere unter Einbeziehung von Nachhaltigkeits-Aspekten).

Vorschlag Themenbereiche:

4. Entwicklung von Systemen zur Automatisierung und Verwaltung von Zugriffsrechten.

Mögliche Förderschwerpunkte umfassen beispielsweise folgenden Bereiche:

- Systeme zur Verwaltung von spezifische (Personen- und Maschinen-) Zugriffsrechten auf Daten und Metadaten (von “Private Data” bis “Open Access”) Die Prinzipien werden in NFDI entwickelt, die technische Verwaltung ist aber integrativer Bestandteil der Data- und Workload-Management-Systeme und deshalb technisch in deren Entwicklung einzuschließen.
- Vereinheitlichung und Homogenisierung der Authentifizierung und der Zugriffsrechte auf Forschungsdaten (innerhalb und zwischen den ErUM-Communities).

Generelle Punkte zur Diskussion:

- Sollte ein Verbundpartner von einer “Research Infrastructures” verpflichtend sein?
- Wie groß sollen die Verbände sein?
- Wie viele Partner von wie vielen verschiedenen Komitees sind verpflichtend?