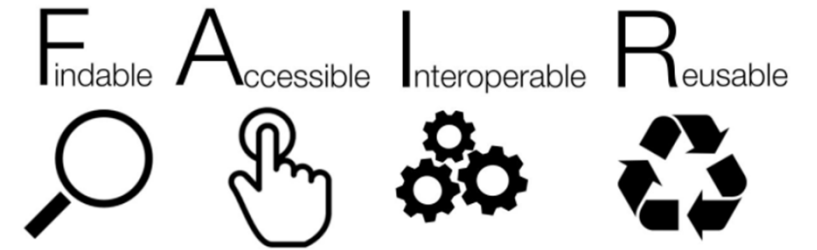


Machine Learning und Research Data Management

Einfluss auf die Teilchen-Förderperiode 2024-27

Martin Erdmann mit Diskussionsbeiträgen von Markus Elsing, Gregor Kasieczka,
Günter Quast, Markus Schumacher

19-Nov-2022



Research Data Management (RDM) (data-life-cycle, FAIR, Portale):

Stakeholder

1. International sind große Konsortien WLCG/CERN, DESY, GSI sehr aktiv, z.B. etablierte RDM Middleware für 1EByte Daten.
2. Punch4NFDI (bereits gefördert), Pionierprojekt wie wir in Zukunft individuelle Datenanalysen entwickeln wollen („science data platform“).
3. ErUM-Data/DIG-UM zielen auf Community-übergreifende Strategien. Gemeinsame Schwerpunkte NFDI/ErUM-Data werden jeweils abgesprochen: Erste Beispiele sind Authentifizierungsfragen und Digital Object Identifiers.

Einfluss auf die Teilchen-Förderperiode 2024-27: Bei Maßnahmen und Entwicklungen zu RDM für große Experimente setzen wir momentan auf Lösungen der großen Konsortien, Punch4NFDI, ErUM-Data (außerhalb ErUM-Pro-Förderung). Kleinere Experimente benötigen eventuell ErUM-Pro-Förderung der antragstellenden Institute.

Methoden modernen maschinellen Lernens (ML) haben sich als „pathbreaking“ etabliert, manche Ansätze werden erst durch ML möglich. Erfolgreiche ML-Verfahren verkürzen Messzeiten und können damit Erkenntnisgewinn beschleunigen (Argumente im Sinn von Sustainability).

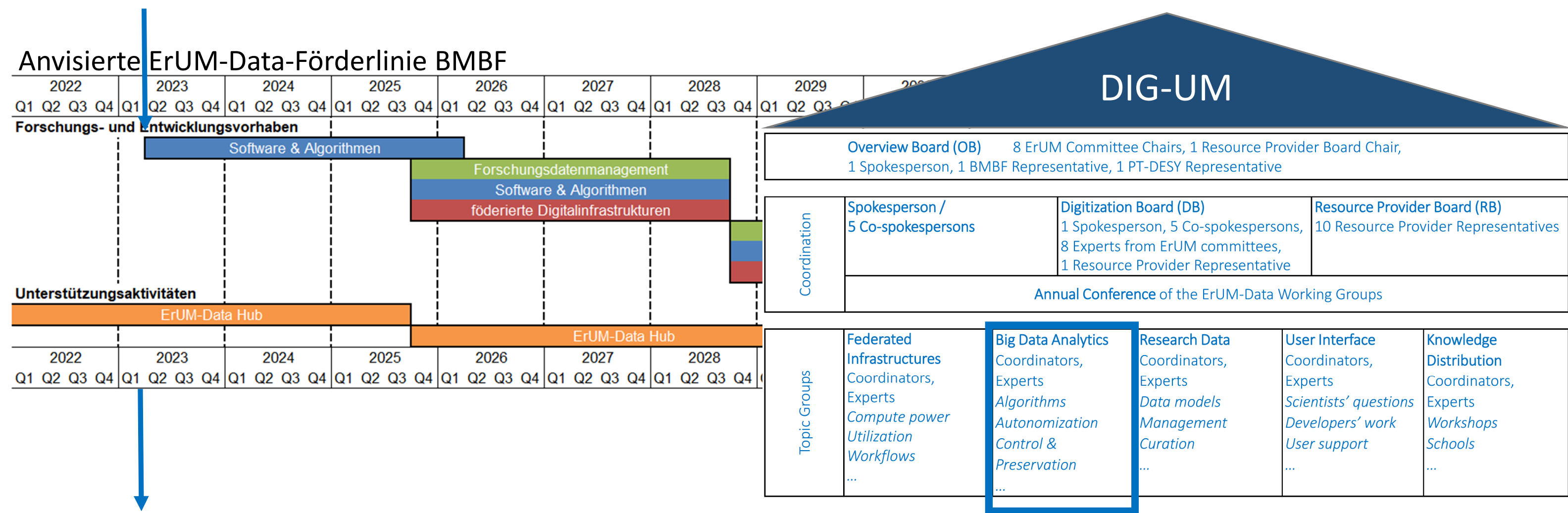
Breites Einsatzgebiet von ML bei deutschen Gruppen: Online-Bereich (FPGA), Spurrekonstruktion, Flavour Tag, Generatives Modelling, Klassifizierung von Kollisionseignissen, Anomalie-Suchen, Datengetriebene Korrekturen, Unfolding/Inferenz etc.
Es gibt Ausbildungsinitiativen GridKa-Schule, Terascale, DIG-UM/ErUM-Data-Hub,...

Internationale Sichtbarkeit der deutschen Beiträge halten, in den Neuentwicklungen weiterhin vorne mit dabei sein und die vielen Vorteile der neuen Technologien ausschöpfen.

Experimentspezifische ML-Entwicklungen zu HEP sind bislang vorwiegend in den antragstellenden Instituten angesiedelt und schließen an institutsspezifische Zielsetzungen an (z.B. CMS-HGCAL) → **Förderanträge im Rahmen von ErUM-Pro.**

Hoffnung *nicht* erfüllt: „Digitale Transformation“ in KET wird durch **ErUM-Data** abgedeckt, „schont“ ErUM-Pro-Fördermittel

Bewilligung 15% von 52 Anträgen, d.h. auch wegweisende ML-Projekte werden nicht gefördert, die wir domain-spezifisch benötigen & weiterhin internationale Sichtbarkeit bedeuten.



23.-24.2.2023 Big Data Analytics Connection Workshop @Univ. Hamburg/DESY: Planung der Bedarfe mit allen Interessierten, geförderten, nicht-geförderten PI's.

→ (Experimentübergreifende) Förderanträge im Rahmen von ErUM-Pro sehr wahrscheinlich.

ERUM-DATA-HUB EVENT PROGRAMME 2023

| | | |
|--------|--------------------------|--------------|
| 23 FEB | BIG DATA ANALYTICS | |
| - | | |
| 24 FEB | CONNECTION WORKSHOP | DESY HAMBURG |
| 27 FEB | DEEP LEARNING SCHOOL | |
| - | | |
| 3 MAR | BASIC CONCEPTS | MEINERZHAGEN |
| 30 MAY | ERUM-DATA SUSTAINABILITY | |
| - | | |
| 2 JUN | EXPERT WORKSHOP | MEINERZHAGEN |
| 19 JUN | ADVANCED DEEP LEARNING | |
| - | | |
| 20 JUN | TRAIN-THE-TRAINER | TU DORTMUND |
| 7 AUG | DEEP LEARNING SCHOOL | |
| - | | |
| 11 AUG | BASIC CONCEPTS | MEINERZHAGEN |
| 25 SEP | ADVANCED DEEP LEARNING | |
| - | | |
| 29 SEP | ACTIVE TRAINING COURSE | MEINERZHAGEN |

MORE TO BE ANNOUNCED SOON:

Research Data Management (**Location TBA**)
Train-the-Trainer Deep Learning Basics (**München**)
Next Generation Environment for Interoperable Data Analysis (**Berlin**)

Big Data Analytics Connection Workshop @Univ. Hamburg/DESY
<https://indico.desy.de/event/36573/>

Bitte machen Sie Ihre Masterand:innen & Doktorand:innen auf die Intensivkurse zu Deep Learning aufmerksam.

Pfingstwoche: Ziel ist Write-up „Sustainability in ErUM“. Federated Infrastructures, Big Data Analytics, Research Data Management
Gesucht: Wer hilft bei Programmplanung? (ErUM-Data-Hub organisiert)

www.erumdatahub.de

