

GeRDI

Generic Research Data Infrastructure

Richard Grunzke

richard.grunzke@tu-dresden.de

Challenge

- Heterogeneous data infrastructures
- Various disciplines
- Data sources and formats, metadata schemas and stores, repositories, archives, data management systems, ...
- Data (pre)processings, HPC, workflows, Big Data frameworks, ...
- Security, AAI, licenses, ...
- Access via commandline, desktop, and web clients, APIs, ...

Overall Aims

- Development of a Generic Research Data Infrastructure (GeRDI)
- Connect arbitrary research data systems to enable searching
- Blueprint with reference implementation for an individual system
- Focus on long tail research data at (mainly) universities
- Three pilot centers in first phase
- Germany-wide offering in second phase

Project overview

- 2.5M Euro DFG funding
- Start at 11/2016
- 3 year duration
- 5 partners
 - Deutsche Zentralbibliothek für Wirtschaftswissenschaften (ZBW)
 - Deutsches Forschungsnetz (DFN)
 - Technische Universität Dresden (TUD)
 - Leibniz-Rechenzentrum (LRZ)
 - Christian-Albrechts-Universität zu Kiel (CAU)
- 12 full time equivalents, 2 at ZIH
- 5 work packages

Work Packages

AP 1: Anforderungen und Software

Leitung: Universität Kiel

TP1.1
Fallstudien (ZBW)

TP1.2
Analyse, Spezifikation
(Uni Kiel)

TP1.3
Softwaremanagement
(Uni Kiel)

AP 2: Umsetzung der förderierten Infrastruktur

Leitung: TU Dresden

TP2.1
Software- und System-
Architektur (Uni Kiel)

TP2.2
Metadaten (ZBW)

TP2.3
Metadatenmanagement
(TUD)

TP2.4
Datenmanagement
(TUD)

AP 3: Pilotbetrieb

Leitung: LRZ München

TP3.1
IT-Sicherheit (LRZ)

TP3.2
Aufbau, Integration der
Pilotzentren (TUD)

TP3.3
Community
Management und
Datenkuration (ZBW)

TP3.4
Evaluierung, Proof of
Concept (TUD)

AP 4: Nachhaltigkeit

Leitung: ZBW

TP4.1
Trainingskonzept (DFN)

TP4.2
Zukünftiges
Betriebsmodell (DFN)

TP4.3
Vorbereitung Roll-Out
(ZBW)

AP 5: Governance, Projektmanagement und Dissemination

Leitung: K. Tochtermann (ZBW), H.-J. Bungartz (DFN)

Requirements and Software

- CAU leading
- Creation of case studies
- Analysis of technical and user requirements
- Software management

Implementation of Federated Infrastructure

- TUD leading
- Software and system architecture
- Metadaten
- Metadatenmanagement
- Datenmanagement

Pilot Operation

- LRZ leading
- IT Security
- Build up of pilot centers
- Community management and data curation
 - Ocean sciences (CAU)
 - Hydrology, meteorology, climate (LZR)
 - Life sciences, digital humanities (TUD)
- Evaluation and proof-of-concept

Sustainability

- ZBW leading
- Training concept
- Future operation model
- Roll-out preparation

Governance, Project Management and Dissemination

- Tochtermann (ZBW) and Bungartz (LZR/DFN) leading
- Governance
 - Executive committee
 - User committee
 - Advisory board
- Project management
- Dissemination

Conclusion

- Pilot phase
- Connect research data systems to enable search
- Blueprint with reference implementation
- Focus on long tail research data
- Extended offering in second phase

Thank you
