



Universität Stuttgart
Universitätsbibliothek



Bibliotheksservice-Zentrum
Baden-Württemberg



DA-FDM

Digitaler Assistent für das
Forschungsdatenmanagement

Erweiterung des DA-3 zur
Sacherschließung von
Forschungsdaten

Imma Hinrichs
Karoline Weinspach

13.11.2024



Baden-Württemberg.de

Gefördert vom Ministerium
für Wissenschaft, Forschung und Kunst Baden-Württemberg

Projekt DA-FDM

Projektidee:

- Sichtbarkeit und Auffindbarkeit von Forschungsdaten verbessern
- Unterstützung der Forschenden bei der inhaltlichen Beschreibung ihrer Daten mit eingeführten Vokabularien
- Nachweis der Forschungsdaten in Katalogen mit inhaltlicher Erschließung

- **Projektpartner**



Universität Stuttgart
Universitätsbibliothek



- **Förderung** durch das MWK BW



Baden-Württemberg.de

- Projektstart 1. April 2023, Laufzeit: 2 Jahre

Forschungsdatenmanagement - FDM

- Umfasst alles von Planung, Erhebung und Analyse bis Speicherung, Archivierung und Veröffentlichung
- Reine Forschungsdatensätze oder Addenda zu Publikationen
- Eigene Repositorien
- FAIR Data Principles (<https://www.go-fair.org/fair-principles/>)
 - Findability
 - Accessibility
 - Interoperability
 - Reusability

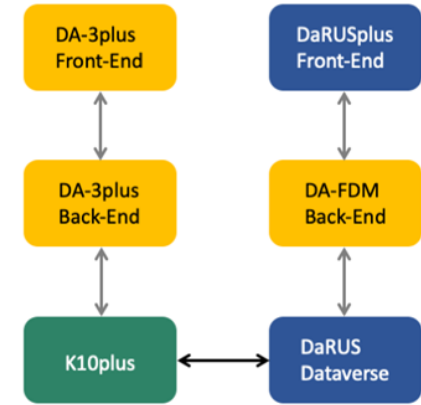
Projekt DA-FDM

Umsetzung des Projekts

- Training einer KI, die Vorschläge aus Vokabularen generiert
- Entwicklung Schnittstellen Repositorien – Verbundkatalog
- Verwendung nachnutzbarer Standards für Protokolle und Schnittstellen

Rahmenbedingungen

- Entwicklung des **DA-FDM** für das **DatenRepository** der **Universität Stuttgart** und die Repositoriums-Software **Dataverse®**
- Universität Stuttgart → Fokus auf Ingenieurwissenschaften
- Nachweis im Verbundkatalog **K10plus**, Nutzung des  **DA-3** für weitere Erschließung



DA-FDM

Digitaler Assistent für das Forschungsdatenmanagement

Herausforderungen

- Unterschiedliche Anforderungen der verschiedenen involvierten Parteien
- Geeignete Ebenen für Erschließung
- Geeignetes Erschließungssystem
- Geeignete Schnittstellen zwischen Repositorien und Katalogen
- Geeignete Menge an Trainingsdaten
- Nachnutzbarkeit

Datensätze werden mit Metadaten versehen wie

- Title, Author, Description, Subjects, Keywords, ...
- Link zu Related Publication, Funding Information (DFG, BMBF, ...)

Problematik freier Keywords

- Mathematics – Maths
- Biochemistry – Biological Chemistry
- Physics – Physisc

The screenshot displays the DaRUS website interface. At the top, the University of Stuttgart logo and 'Germany' are visible. A navigation bar includes 'Search', 'About', 'User Guide', 'Support', and 'Log In'. The main content area shows the dataset title and a brief description. A sidebar on the left provides additional context, including the author 'Peters, Rene' and the dataset version 'Version 1.0'. The right sidebar contains a detailed metadata table.

Field	Value
Persistent Identifier	doi:10.18419/darus-3614
Publication Date	2023-09-14
Title	Replication Data for: Cooperative Lewis Acid-1,2,3-Triazolium-Aryloxide Catalysis: Pyrazolone Addition to Nitroolefins as Entry to Diaminoamides
Author	Peters, Rene (Universität Stuttgart) - ORCID: 0000-0002-6668-4017
Point of Contact	Use email button above to contact. Peters, Rene (Universität Stuttgart)
Description	development and optimization of catalysis, mechanistic experiments, synthetic description of ligand and catalyst synthesis, application of catalysts, synthetic use of products, spectroscopic characterization data of all new products, EPR data of complexes, X-ray data, biological activity of catalysis products and derivatives thereof, for interpretation see manuscript (2023-08-30)
Subject	Chemistry
Keyword	Catalyst Development (TEMA) https://purl.org/tema/113291 Asymmetric Catalysis (TEMA) https://purl.org/tema/113538 Reaction Mechanism (TEMA) https://purl.org/tema/017811 Organic Synthesis (TEMA) https://purl.org/tema/131654 EPR (Wikidata) https://www.wikidata.org/wiki/Q260463 X-ray Crystal Structure Analysis Cell Painting Assay
Related Publication	Wanner, D. M., Becker, P. M., Suhr, S., Wannenmacher, N., Ziegler, S., Hermann, J., Willig, F., Gabler, J., Jangid, K., Schmid, J., Hans, A. C., Frey, W., Sarkar, B., Kästner, J., Peters, R., Cooperative Lewis Acid-1,2,3-Triazolium-Aryloxide Catalysis: Pyrazolone Addition to Nitroolefins as Entry to Diaminoamides, Angew. Chem. Int. Ed. 2023, 62, e202307317. doi: 10.1002/anie.202307317 https://doi.org/10.1002/anie.202307317
Funding Information	DFG: 358283783 - SFB 1333 DFG: 310990693 DFG: INST 40/575-1 FUGG (JUSTUS 2 cluster) Ministry of Science, Research and Arts Baden-Württemberg: bwHPC European Union: Drug Discovery Hub Dortmund (DDHD), EFRE-0200481 European Union: Innovative Medicines Initiative (grant agreement number 115489)
Project	Cooperative Asymmetric Dual / Multiple Activation Catalysis Under Confinement (Level 0)
Depositor	Peters, Rene
Deposit Date	2023-07-24

<https://darus.uni-stuttgart.de/dataverse/darus>

DA-FDM

Digitaler Assistent für das Forschungsdatenmanagement

Mögliche Erschließungsebenen

- 3 Ebenen in DaRUS
 - Subject: Festgelegte grobe Fachauswahl von Dataverse®
 - Keyword: Höchster Detailgrad
 - Topic Classification: Mittlere Erschließungstiefe
- Topic & Keyword: Bisher freie Stichwörter oder Begriffe aus Vokabularen inkl. Link (letzteres meist hinzugefügt von DaRUS-Team)



Subject ⓘ	Physics
Keyword ⓘ	Continuous phase transition (PhySH) https://doi.org/10.29172/6a6a8f Directed Percolation (PhySH) https://doi.org/10.29172/def3e8c1-b28b Continuous percolation transition (PhySH) https://doi.org/10.29172/b4 Entanglement Entropy (PhySH) https://doi.org/10.29172/4a7f1f86-798 Entanglement Production (PhySH) https://doi.org/10.29172/95477339 Phase transitions (PhySH) https://doi.org/10.29172/f2a9db78-1651-4a Quantum circuits (PhySH) https://doi.org/10.29172/27d61d4c-d801-47 Quantum Entanglement (PhySH) https://doi.org/10.29172/5de0faa9-6f Spin Lattice Models (PhySH) https://doi.org/10.29172/6ff84248-428b-4 Monte Carlo Methods (PhySH) https://doi.org/10.29172/eb9bd2e1-ee Critical Exponents (PhySH) https://doi.org/10.29172/5c05a702-8101-4 Numerical Techniques (PhySH) https://doi.org/10.29172/45a22032-e2
Topic Classification ⓘ	Theoretical Condensed Matter Physics (DFGFO) https://github.com/tit
Related Publication ⓘ	Makki, Nastasia, Lang, Nicolai & Büchler, Hans Peter (2024). Absorbin Res. 6, 013278. doi: 10.1103/PhysRevResearch.6.013278 ⓘ https:// Makki, N., Lang, N. & Büchler, H. P. (2023). Absorbing State Phase Tr

Zunächst Fokus auf mittlere Ebene (Topic), im nächsten Schritt Auswahl von fachspezifischen Vokabularen (Keyword)

DA-FDM

Digitaler Assistent für das Forschungsdatenmanagement

Anforderungen an Erschließungssysteme

- Muss auch neue Forschungsgebiete abdecken
- Muss mit wachsender Menge von Forschungsdaten skalieren
- Mehrsprachig (min. EN und DE)
- Trainingsdaten für (halb-)automatische Erschließung müssen vorhanden sein
- Geeignete Katalogstrukturen vorhanden
- Knowhow bei Forschenden und Fachpersonal vorhanden
- Langfristige Pflege sichergestellt
- Frei verfügbar, keine Lizenzabgaben

DA-FDM

Digitaler Assistent für das Forschungsdatenmanagement

Auswahl eines geeigneten Erschließungssystems (Topic-Level)

- Bibliothekarische Klassifikationssysteme wurden nicht für FD gemacht
 - Aktueller Stand der Forschung nicht abgebildet
 - Hierarchische Ordnung
 - Thema und Methode schwer abgrenzbar
- Spezielle an FD angepasste Systeme entstehen derzeit, können später verwendet werden



Die **DFG-Systematik** wurde insbesondere aufgrund ihrer hohen Aktualität durch ständige Anpassung (alle vier Jahre) sowie ihrer Bekanntheit unter Forschenden gewählt

Digitaler Assistent für das Forschungsdatenmanagement

- Liegt auf deutsch und englisch vor
- Wird alle 4 Jahre aktualisiert
- Wissenschaftsbereiche
 - Geistes- und Sozialwissenschaften
 - Lebenswissenschaften
 - Naturwissenschaften
 - Ingenieurwissenschaften

- 15 Fachgebiete, 49 Fachkollegien, 214 Fächer
- Im Projekt werden die beiden Ebenen Fachkollegium und Fach verwendet

Wissenschaftsbereich	Fachgebiet	Fachkollegium	Fach
Orange	Orange	Orange	Orange
Red	Red	Red	Red
Green	Green	Green	Green
Blue	Blue	Blue	Blue

<https://www.dfg.de/de/foerderung/antrag-foerderprozess/interdisziplinaritaet/faecherstruktur>

- Ausschnitt aus Systematik 2020-2024

405-06	Polymere und biogene Werkstoffe und darauf basierende Verbundwerkstoffe			
406-01	Herstellung und Eigenschaften von Funktionsmaterialien			
406-02	Biomaterialien			
406-03	Thermodynamik und Kinetik sowie Eigenschaften der Phasen und Gefüge von Werkstoffen	406		4
406-04	Computergestütztes Werkstoffdesign und Simulation von Werkstoffverhalten von atomistischer bis mikroskopischer Skala	Materialwissenschaft		Ingenieurwissenschaften
407-01	Automatisierungstechnik, Regelungssysteme, Robotik, Mechatronik, Cyber Physical Systems			
407-02	Messsysteme			
407-03	Mikrosysteme	407		
407-04	Verkehrs- und Transportsysteme, Intelligenter und automatisierter Verkehr	Systemtechnik		
407-05	Arbeitswissenschaft, Ergonomie, Mensch-Maschine-Systeme			
407-06	Biomedizinische Systemtechnik			
408-01	Elektronische Halbleiter, Bauelemente und Schaltungen, Integrierte Systeme			
408-02	Nachrichten- und Hochfrequenztechnik, Kommunikationstechnik und -netze, Theoretische Elektrotechnik	408		44
408-03	Elektrische Energieerzeugung, -übertragung, -verteilung und -anwendung	Elektrotechnik und Informationstechnik		Informatik, System- und Elektrotechnik
409-01	Theoretische Informatik			
409-02	Softwaretechnik und Programmiersprachen			
409-03	Sicherheit und Verlässlichkeit			
409-04	Rechner, Kommunikation, Datenbank und verteilte Systeme			

<https://www.dfg.de/de/foerderung/antrag-foerderprozess/interdisziplinaer/faecherstruktur>

DA-FDM

Digitaler Assistent für das Forschungsdatenmanagement

Schnittstelle DaRUS – K10plus

- Datensätze werden aus DaRUS über OAI-Schnittstelle im DataCite-Format exportiert
- Vor Übernahme in K10plus ist eine Konvertierung der Datensätze aus DaRUS in das PICA-Format (Datenformat des K10plus) notwendig
- Dazu wurde ein Mapping von DataCite nach PICA erstellt
- Mapping dient als Grundlage für den Konverter, der die Daten von DataCite nach PICA umsetzt und per Skript in den K10plus einspielt



A01

SFB-TRR 161 A01 "Uncertainty Quantification and Analysis in Visual Computing"

(Universität Stuttgart, Universität Konstanz)

DaRUS > SFB/Transregio 161 "Quantitative Methods for Visual Computing" > SFB-TRR 161 A01 "Uncertainty Quantification and Analysis in Visual Computing" >

Supplemental Material for Out-of-Core Dimensionality Reduction for Large Data via Out-of-Sample Extensions

Version 1.0

Supplemental Material
for Out-of-Core
Dimensionality
Reduction for Large
Data via Out-of-
Sample Extensions

Reichmann, Luca; Hägele, David; Weiskopf, Daniel, 2024, "Supplemental Material for Out-of-Core Dimensionality Reduction for Large Data via Out-of-Sample Extensions", <https://doi.org/10.18419/darus-s-4441>, DaRUS, V1, UNF:6:WoQ4Mnftz92VcvZ/qCGL5w== [fileUNF]

[Cite Dataset ▾](#)

Learn about [Data Citation Standards](#).

Access Dataset ▾

Contact Owner

Share

Dataset Metrics ⓘ

686 Downloads ⓘ

Description ⓘ

This dataset contains the supplemental material for "Out-of-Core Dimensionality Reduction for Large Data via Out-of-Sample Extensions".

The contents and usage of this dataset are described in the `README.md` files.

Subject ⓘ

Computer and Information Science

Keyword ⓘ

Information Visualization, Dimension Reduction (Statistics)

Related Publication ⓘ

L. Reichmann, D. Hägele, D. Weiskopf, "Out-of-Core Dimensionality Reduction for Large Data via Out-of-Sample Extensions" in arXiv preprint, 2024. [arXiv: 2408.04129](#)

Notes ⓘ

Part of this dataset is a large collection of images (*.png) and matrix files (*.npz) located in the `usecase` directory, which can be difficult to download via the web interface. There are scripts to automatically download the files via command line interface, `download_images.py` and `download_embeddings.py`.

The scripts can be execute with Python3, e.g. `python3 download_images.py`.

License/Data Use Agreement

[Custom Dataset Terms](#)

Eingabe: S1UB:05-09-24 Änderung: S1UB:09-09-24 11:42:31 Status: S1UB:05-09-24

0100 1902801784

0500 Oax

0501 Computerdaten\$bcod

0502 Computermedien\$bc

0503 Online-Ressource\$bcrc

0575 darus

1100 2024

1131 [1857755366](#)!Forschungsdaten ; ID: gnd/1098579690

1500 eng

2051 10.18419/darus-4441

2240 KXP: 1902801784

3000 Reichmann, Luca\$iorcid\$0009-0002-1041-2693\$BVerfasserIn\$4aut

3010 Hägele, David\$iorcid\$0000-0002-2679-6882\$BVerfasserIn\$4aut

3010 Weiskopf, Daniel\$iorcid\$0000-0003-1174-1026\$BVerfasserIn\$4aut

4000 Supplemental Material for Out-of-Core Dimensionality Reduction for Large Data via Out-of-Sample Extensions

4020 Version: 1.0

4030 Stuttgart\$Universität Stuttgart

4060 1 Online-Ressource (2018 Dateien)

4201 Updated: 2024-09-02

4207 This dataset contains the supplemental material for "Out-of-Core Dimensionality Reduction for Large Data via Out-of-Sample Extensions". The contents and usage of this dataset are described in the README.md files.

4207 Part of this dataset is a large collection of images (*.png) and matrix files (*.npz) located in the usecase directory, which can be difficult to download via the web interface. There are scripts to automatically download the files. To be executed with Python3, e.g. python3 download_images.py.

4950 <https://doi.org/10.18419/darus-4441>\$xR\$4LF

4985 Open Access

5060 [dfg]4.43-05\$Image and Language Processing, Computer Graphics and Visualisation, Human Computer Interaction, Ubiquitous and Wearable Computing\$ADE-93\$ADaRUS

5060 [dfg]4.43-05\$Bild- und Sprachverarbeitung, Computergraphik und Visualisierung, Human Computer Interaction, Ubiquitous und Wearable Computing\$ADE-93\$ADaRUS

5520 [FS]Computer and Information Science\$ADE-93\$ADaRUS

5520 Information Visualization\$0<https://id.loc.gov/authorities/subjects/sh2002000243.html>\$ADE-93\$ADaRUS

5520 Dimension Reduction (Statistics)\$0<https://id.loc.gov/authorities/subjects/sh2010000188.html>\$ADE-93\$ADaRUS

E001 05-09-24 : I01

4801 Elektronische Ressource - Zugang über das Repositorium

7100 \$B93\$Dn

7133 <https://doi.org/10.18419/darus-4441>

7901 S1UB:05-09-24

7902 05-09-24

7903 05-09-24 11:01:45.000

8012 daru

7800 4567299388

DA-FDM

Digitaler Assistent für das Forschungsdatenmanagement

Trainingsdaten

- Benötigt werden Datensätze mit
 - zugehörigen Publikationen (Best Case: Open Access)
 - zugeordneter Erschließung
- Zugehörige Publikationen häufig nicht frei verfügbar, dann nur Zusammenfassung verwendbar
- Erschließung von Trainingsdatensätzen im Projekt
- Mit erschlossenen Datensätzen wird KI trainiert



DA-FDM

Digitaler Assistent für das Forschungsdatenmanagement

Aktueller Stand des Projekts

- Digitaler Assistent wird laufend mit Trainingsdaten gefüttert und Ergebnisse werden analysiert
- DaRUS-K10plus-Schnittstelle
 - Testeinspielung aller Forschungsdatensätze aus DaRUS (ca. 1650 Datensätze)
 - Prüfung durch UB Stuttgart und BSZ
 - Erzeugung von Exemplarsätzen der UB Stuttgart
- Zusammenarbeit mit TIB zur Weiterentwicklung der DFG-Systematik als Vokabular

DA-FDM

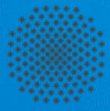
Digitaler Assistent für das Forschungsdatenmanagement

Nächste Schritte

- Tests Prototyp DA-FDM inkl. Schnittstellen
- Erste produktive Einspielung in den K10plus
- Konzeption eines Workflows für Updates in DaRUS
 - Z.B.: bei Ergänzung von Sacherschließung mit dem DA-FDM
- Konfiguration einer Routine zur Abholung von Updates im K10plus (Schnittstelle K10plus-DaRUS)

Community

- Trainingsdaten / Datensätze mit zugehörigen Publikationen?
- Finaler Workshop zum Ende des Projekts



Universität Stuttgart
Universitätsbibliothek

Vielen Dank!



Dr. Imma Hinrichs – Koordination Sacherschließung
Karoline Weinspach – Projektleitung DA-FDM

E-Mail imma.hinrichs@ub.uni-stuttgart.de
karoline.weinspach@ub.uni-stuttgart.de

Universitätsbibliothek Stuttgart
Holzgartenstr. 16
70174 Stuttgart

<https://www.ub.uni-stuttgart.de>