



NFDI4ing

Konferenz 2021

27./28.9. online

Herzlich Willkommen zur Konferenz der Nationalen Forschungsdateninfrastruktur für die Ingenieurwissenschaften 2021.

Die diesjährige Konferenz, ausgerichtet von der Technischen Universität Braunschweig, in Kooperation mit der Technischen Universität Dresden und der Sächsischen Landes- und Universitätsbibliothek findet dieses Jahr online statt. In parallelen Sessions zu den Themengebieten „Infrastruktur“, „Forschungssoftware“, „Best-Practices“ und „Kooperation“ können Sie sich über aktuelle Entwicklungen im Forschungsdatenmanagement in den Ingenieurwissenschaften informieren. In Workshops haben Sie zudem die Möglichkeit für weiteren Austausch und Mitgestaltung.

Für eine Teilnahme bitten wir Sie sich über die Webseite nfdi4ing.de/konferenz anzumelden. Die Onlinekonferenz wird über die Plattform Zoom stattfinden, für welche Ihnen einige Tage vor Beginn die Login-Daten zugesandt werden. Wir empfehlen die Installation des Zoom-Client auf Ihrem Rechner. Technische Hilfestellungen finden sich in einem separaten Dokument auf unserer Webseite. Nachfolgend finden Sie den Kurzüberblick über das Programm sowie im Anhang kurze Inhaltsbeschreibungen der Beiträge. Für deren Inhalt sind jeweils die Referenten verantwortlich. Bei weiteren Fragen nutzen Sie ebenfalls gern das Kontaktformular auf unserer Webseite.

Die Referenten freuen sich auf den Austausch mit Ihnen und wünschen Ihnen viele Erkenntnisse zur Zukunft des Forschungsdatenmanagements!

Welcome to the National Research Data Infrastructure for Engineering 2021 Conference.

This year's conference, hosted by the Technische Universität Braunschweig (TUB), in cooperation with Technische Universität Dresden (TUD) and the Saxon State and University Library (SLUB), will take place online this year. In parallel sessions on the topics of "Infrastructure", "Research Software", "Best Practices" and "Cooperation", you can find out about current developments in research data management in the engineering sciences. In workshops you will also have the opportunity for further exchange and to shape the future of research data management.

To participate, please register via the website nfdi4ing.de/conference. The online conference will take place via the Zoom platform, for which the login data will be sent to you a few days before the start. We recommend installing the Zoom client on your computer. Technical assistance can be found in a separate document on our website.

Below you will find a brief overview of the programme as well as short descriptions of the contributions in the appendix. The speakers are solely responsible for their content. If you have any further questions, please feel free to use the contact form on our website.

The speakers look forward to exchanging ideas with you and wish you many insights into the future of research data management!

Hinweise / Notes

Zeichenerklärung / Symbols



Workshop



English presentation



English presentation possible if necessary
(slides possibly in German)

Themengebiete / Topics



Kooperation



Forschungssoftware



Best-Practices



Infrastruktur

Übersicht der mit Beiträgen vertretenen Institutionen / Contributing Institutions

BAFG	Bundesanstalt für Gewässerkunde	TIB	Technische Informationsbibliothek Hannover
DLR	Deutsches Zentrum für Luft- und Raumfahrt	TH Koeln	Technische Hochschule Koeln
FIT	Fraunhofer-Institut für Angewandte Informationstechnik FIT	TUBS	Technische Universität Braunschweig
FZ Juelich	Forschungszentrum Juelich	TUC	TU Clausthal
HRLS	Hochleistungsrechenzentrum Stuttgart	TUD	Technische Universität Dresden
HS Karlsruhe	Hochschule Karlsruhe	TUDA	Technische Universität Darmstadt
	Jacobs University Bremen	TUM	Technische Universität München
KIT	Karlsruher Institut für Technologie	UCL	University College London
LUH	Leibniz Universität Hannover		Univ. Duisburg-Essen
MPSD	Max-Planck Institute for Structure and Dynamics of Matter		Universität Münster
RWTH	RWTH Aachen		Universität Stuttgart
SLUB	Sächsische Landes- und Universitätsbibliothek Dresden		

Agenda / agenda

Tag 1 – 27.09.2021

9:00 – 10:15 Eröffnung

Begrüßung
Grußworte
Keynote
Vorstellung der Agenda / Hinweise

Manfred Krafczyk (TUBS)
York Sure-Vetter (NFDI)
Robert Schmitt (RWTH)
Jan Linxweiler (TUBS)

10:15 – 10:30 Pause

10:30 – 12:00 Sessions A

A1	Beschreibung ingenieurwissenschaftlicher Forschungsdaten - Workshop rund um Kompetenzfragen zur Anforderungsermittlung	Dorothea Iglezakis (Univ. Stuttgart) Sophia Leimer (Univ. Duisburg-Essen) Susanne Arndt (TIB)	 
A2	Nachnutzung von Forschungssoftware in komplexen Softwareworkflows	Patrick Kuckertz (FZ Juelich) Oliver Karras (TIB) Felix Engel (TIB) Jan Göpfert (FZ Juelich)	 
A3	Elektronische Laborbücher und ihr Einsatz in der Lehre - Beispiele aus ingenieurwissenschaftlichen Praktika	Florian Strauß (TUC)	
	Datenkompetenz von Anfang an - Bausteine im Ingenieurstudium	Manuela Richter (TUDA)	
	Bildung, Aus- und Weiterbildung im Forschungsdatenmanagement - Angebote für die Ingenieurwissenschaften	Amelie Metzmacher (RWTH)	

Moderation: Évariste Demand

12:00 – 13:00 Mittagspause

13:00 – 14:30 Sessions B

B1	Die FAIR-Prinzipien und Vertrauensschutz im HPMC-Bereich –Schwierigkeiten und Lösungsansätze	Benjamin Farnbacher (TUM) Martin Thomas Horsch (HRLS) Christian Stemmer (TUM) Manuel Hora (TUM) Nils Hoppe (TUM)	
	Research software development with Julia	Michael Schlotke-Lakemper (Univ. zu Köln) Hendrik Ranocha (Univ. Münster)	
B2	H5Flow - Data Exploration and Sharing using HDF5 for Fluid Mechanics Problems	Matthias Probst (KIT)	
	Datacubes: Data Organization, Access, and Analytics (also) for Science and Engineering	Peter Baumann (Jacobs University)	

Moderation: Frank Löffler

B3

Produktionsdaten Management von der Edge in die Hochschul-Cloud

Moritz Kröger (RWTH)
Michael Rath (RWTH)

Gelebtes Forschungsdatenmanagement in den Materialwissenschaften: Workflows zur Weiterentwicklung von Point-of-Care-Antigen-Tests

Patrick Altschuh (HS Karlsruhe)

KaDI4Mat - Forschungsdateninfrastruktur für die Materialforschung

Michael Selzer (KIT)

*Moderation: David Hecker***B4**

Test Driven Development (TDD)

Sören Peters (TUBS)
Sven Marcus (TUBS)**14:30 – 14:45** **Pause****14:45 – 16:00** **Journal Vorstellung & Award**

ing.grid – Data Management in Engineering Sciences

Peter Pelz

Lightning Talks zu nominierten FDM-Lösungen (je 5')

verschiedene

Verleihung NFDI4Ing Award

Jan Linxweiler

Tag 2 – 28.09.2021**08:50 – 09:00** **Begrüßung****09:00 – 10:30** **Sessions C****C1**

Strukturierte Metadaten in den Verkehrswissenschaften

Matthias Fuchs (SLUB)
Susanne Arndt (TIB)
Maximilian Bäumlner (TUD)
Stefanie Roski (SLUB)**C2**

Einführung in Git

Dennis Gläser (Univ. Stuttgart)



Energiedaten für die studentische Ausbildung in python-Energiedatenanalyse gesucht

Johanna May (TH Koeln)

**C3**

Forschungsdatenmanagement in einem Großforschungsverbund: Erfahrungen aus dem Exzellenzcluster POLIS

Ephraim Schoof (KIT)

Anforderungen für Forschungsdatenmanagement und integriertes maschinelles Lernen zwischen Medizin und Ingenieurwissenschaften

Julian Grolig (KIT)

*Moderation: Michael Selzer*

Pace3D - Ein Framework für High Performance Mikrostruktursimulation zur datengetriebene Entwicklung neuer Materialien

Britta Nestler (KIT)

C4

Bembel - The BEM-based engineering library

Felix Wolf (TUDA)



Semantik im Internet of Production: Erfahrungen aus 3 Jahren Exzellenzcluster

Johannes Theissen-Lipp (RWTH)
Iraklis Dimitriadis (FIT)*Moderation: Martin Hock***10:30 – 11:00** **Pause**

11:00 – 12:30 Sessions D

D1	Visualization of networks – analyzing and visualizing connections between (planned) NFDI consortia	Évariste Demandt (RWTH) Lukas Bossert (RWTH) Sama Majidian (RWTH)	 
	Continous Integration	Dennis Gläser (Univ. Stuttgart)	
D2	Continuous Integration of Scientific Software	Tomislav Maric (TUDA)	
	Making software citation easi(er) – The Citation File Format and its integrations	Stephan Druskat (DLR)	
	Erfahrungen mit dem Testen des Gewässergütemodells QSim	Jens Wyrwa (BAFG)	

Moderation: Dennis Gläser

D3	Data-Stewards in Verbundforschung	Peter Pelz (TUDA)	
	Monitoring von Versuchsobjekten in einem Forschungsdatenmanagementsystem am Beispiel einer Tailored-Forming-Prozesskette	Iryna Mozgova (LUH)	
	Validation as a specific aspect of testing the CORSIKA Air Shower Monte Carlo Simulation package	Ralf Ulrich (KIT)	

Moderation: Marco Berger

12:30 – 13:30 Mittagspause

13:30 – 15:00 Sessions E

E1	plotID, Toolkit zur Verknüpfung von Abbildungen und Forschungsdaten	Jan Lemmer (TUDA) Martin Hock (TUDA) Manuela Richter (TUDA)	
	Datencontainer mit Provenienz: Einfacher Datenaustausch mit verteilten Stakeholdern	Frank Dressel (DLR)	
	Ein ontologiebasiertes Datenmanagementsystem fuer die Produktionstechnik	Iraklis Dimitriadis (FIT) Martin Kretschmer (FIT)	

Moderation: Florian Strauß

E2	Testing open source Python-centric research software	Hans Fangohr (MPSD)	
	Continuous testing in HPC fortran: lessons learned from TROVE	Jamie Quinn (UCL) Anastasias Georgoulas (UCL)	
	Continuous struggle: precise, performant and portable code development in HPC	Ivo Kabadshow (FZ Juelich)	

Moderation: Dennis Gläser

E3	Was Forschende wollen - Erfahrungsbericht aus FDM-Beratungen	Andreas von der Dunk (SLUB)	
	Akzeptanz- und Motivationsförderung mit FDM-User Stories	Amelie Metzmacher (RWTH) Tobias Hamann (RWTH) Mario Moser (RWTH)	

15:00 – 15:30 Verabschiedung

Autorenreferate / *abstracts*

Nachfolgend finden sich die Kurzbeschreibungen der Autoren für die jeweiligen Beiträge.

Sessions A

A1

Beschreibung ingenieurwissenschaftlicher Forschungsdaten - Workshop rund um Kompetenzfragen zur Anforderungsermittlung

Dorothea Iglezakis (Univ. Stuttgart)
Sophia Leimer (Univ. Duisburg-Essen)
Susanne Arndt (TIB)

Stellen wir uns vor, in der Zukunft wären Forschungsergebnisse so beschrieben, erschlossen und miteinander verknüpft, dass man beliebige Fragen an sie stellen könnte. Welche Anfragen würden Sie dann an einen so strukturierten und semantisch erfassten Datenbestand stellen wollen? Um diese Vision Wirklichkeit werden zu lassen benötigen wir Datenstrukturen, die sowohl von Menschen, wie auch von Maschinen verarbeitbar sind. Ontologien, die einen Teil der Welt semantisch modellieren, können ein Baustein dafür sein. Welche Konzepte brauchen wir, um ingenieurwissenschaftliche Forschungsergebnisse zu beschreiben? Welche Fragen will man stellen und beantworten können? Um die Ontologie-Entwicklung innerhalb von NFDI4Ing nutzerorientiert zu gestalten, möchten wir in diesem Workshop gemeinsam mit Ihnen diese sogenannten Kompetenzfragen formulieren und diskutieren.

A2

Nachnutzung von Forschungssoftware in komplexen Softwareworkflows

Patrick Kuckertz (FZ Juelich)
Oliver Karras (TIB)
Felix Engel (TIB)
Jan Göpfert (FZ Juelich)

Im Rahmen des Workshops wird die Bedeutung von wissenschaftlicher Software und Softwareworkflows für die Forschung adressiert. Es wird auf die Forschungsfeld-übergreifende Nachnutzung von Softwarekomponenten und insbesondere auf die Prozesse zum Auffinden und Teilen von Software eingegangen. In Zusammenarbeit mit den Teilnehmern werden aktuelle Tools, Plattformen, Standards und Best Practices der Softwarerecherche und -auswahl zusammengetragen, Erfahrungen diskutiert und Handlungsbedarfe identifiziert. Übergeordnetes Ziel ist es Anforderungen abzuleiten, denen im Rahmen der Weiterentwicklung einer auch auf Software ausgerichteten Forschungsdateninfrastruktur Rechnung getragen werden muss.

A3**Elektronische Laborbücher und ihr Einsatz in der Lehre - Beispiele aus ingenieurwissenschaftlichen Praktika** Florian Strauß (TU Clausthal)

Elektronische Laborbücher sind im Laboralltag bereits weit verbreitet und werden im Rahmen der NFDI4Ing auch an verschiedenen Stellen thematisiert. Ihren Einsatz zur Dokumentation der eigenen Forschungsergebnisse auf ingenieurwissenschaftliche Praktika auszuweiten ist somit ein logischer Schritt. Entsprechend naheliegend ist es in einem weiteren Schritt zu betrachten, ob dieses Werkzeug auch eine Verwendung zur Vermittlung der Grundlagen des Forschungsdatenmanagements finden kann. Der Beitrag beschreibt den auf diese Zielsetzung ausgerichteten Einsatz in einem Praktikum zur Hochspannungstechnik inklusive einer im Anschluss durchgeführten Umfrage unter den Studierenden sowie Ansätze in weiteren Praktika mit anderen Schwerpunktsetzungen.

Datenkompetenz von Anfang an - Bausteine im Ingenieurstudium Manuela Richter (TUDA)

FAIRe Daten sind der Rohstoff des 21ten Jahrhunderts. Sie sind Enabler für Transparenz, Nachhaltigkeit, Wissenschaft und Wirtschaft. Der Vermittlung von Datenkompetenz kommt somit eine Schlüsselrolle zu. Der Vortrag zeigt auf wo und wie Datenkompetenz im Bachelorstudiengang Maschinenbau der TU Darmstadt vermittelt wird.

Bildung, Aus- und Weiterbildung im Forschungsdatenmanagement - Angebote für die Ingenieurwissenschaften Amelie Metzmacher (RWTH)

Im Vortrag erfolgt die Vorstellung des Status quo und der nächsten Meilensteine der NFDI4Ing-Aktivitäten zu den Themen Bildung, Aus- und Weiterbildung im Forschungsdatenmanagement (FDM). Hierzu werden erste konkrete und weitere geplante Angebote im Bereich FDM-Training für die ingenieurwissenschaftliche Community präsentiert. Der Vortrag schließt mit einem Blick auf die weiteren Aktivitäten innerhalb des NFDI-Verbundes.

Sessions B**B1****Die FAIR-Prinzipien und Vertrauensschutz im HPMC-Bereich –Schwierigkeiten und Lösungsansätze** Benjamin Farnbacher (TUM), Martin Thomas Horsch (HRLS), Christian Stemmer (TUM), Manuel Hora (TUM), Nils Hoppe (TUM)

Daten aus High Performance Measurement & Computing (HPMC) sind meist nicht mobil und nur lokal gespeichert, weshalb die Einhaltung der FAIR-Prinzipien eine besondere Herausforderung darstellt. Im Konferenzbeitrag werden diese Schwierigkeiten thematisiert und entsprechende Lösungsansätze und Werkzeuge vorgestellt. Um Datensicherheit und -souveränität auf HPMC-Systemen zu gewährleisten, ist es essentiell die entsprechenden Anforderungen und Bedürfnisse der Nutzenden zu kennen. Im einhergehenden Workshop werden deshalb gemeinsam mit der Community mögliche Anforderungen an den Vertrauensschutz für HPMC-Daten erarbeitet. Forschende sollen sich darauf verlassen können, dass Datensicherheit und Datenhoheit für HPMC-Daten dauerhaft gewährleistet sind. Im Rahmen des Workshops werden diese Ansprüche der Community erarbeitet und diskutiert.

B2**Research software development with Julia**Michael Schlotke-Lakemper
(Univ. zu Köln)
Hendrik Ranocha (Univ. Münster)

Julia is a modern, dynamic programming language designed for high-performance scientific computing. It makes it easy to set up collaborative development workflows by providing tools to create reproducible environments, to set up automatic testing, and to share code via registered packages. In this talk, we will give a brief general introduction to Julia and its capabilities, and then focus on aspects that make it an interesting choice for research software developers. Next to a theoretical overview, we will include real-world examples from the development of Trixi.jl, a Julia package for adaptive numerical simulations of fluid flow and other hyperbolic PDEs.

H5Flow - Data Exploration and Sharing using HDF5 for Fluid Mechanics Problems

Matthias Probst (KIT)

Data value decay due to the lack of data management can inhibit the reuse, exchange and traceability of research data. Especially with respect to long-term storage and the data exchange with collaborators, a sustainable data management, which links meta data to raw data within a self-describing file format, is the way to go. Often times, data acquisition of experimental and numerical data can be supported by guidelines and frameworks, that leads to simplified data evaluation. For the field of fluid mechanics one such concept has been implemented and will be presented. It provides a python class called H5Flow, which is a wrapper around an HDF file and provides a framework, which ultimately allows to efficiently explore, visualize, post-process data and share it within collaborations.

Datacubes: Data Organization, Access, and Analytics (also) for Science and Engineering

Peter Baumann (Jacobs University)

Multidimensional discretized data (also known as "rasters", "grids", "arrays" and "datacubes") represent an information category which substantially contributes to today's "Big Data" challenge in science, engineering, business, and more. For example, in the Earth Sciences we find 1D sensor timeseries, 2D satellite imagery, 3D x/y/t image timeseries and x/y/z exploration data, 4D x/y/z/t climate and ocean data. As such, practically all NFDI domains deal with n-D datacubes in some way. While early on OLAP has dominated datacubes in the business field, more recently the advantages of well-structured datacubes and problem-adequate services get recognized. Also standardization has picked up on the issue, and with SQL/MDA ("Multi-Dimensional Arrays") a domain-agnostic extension of SQL has been established; specifically for geo datacubes, the OGC WCPS ("Web Coverage Processing Service") language has been standardized. These languages are practice-proven into the Petabytes and across application domains like astronomy, remote sensing, metocean, human brain analysis, and genetics, to name but a few. A particular contribution datacube engines can make for both scientific and industrial applications is their query flexibility, scalability, and information integration. In our talk we present the state of the art in array services and illustrate it with rasdaman ("raster data manager") which has pioneered Array DBMSs. We outline conceptual model and architecture of rasdaman, discuss application domains investigated, and illustrate how NFDI domains can benefit from a cross-domain datacube infrastructure.

B3**Produktionsdaten Management von der Edge in die Hochschul-Cloud**Moritz Kröger (RWTH)
Michael Rath (RWTH)

Im Excellent Cluster Internet of Production hat die Cluster4aCluster Initiative vor einigen Jahren mit ihrer Arbeit an einem Universitätsweiten Kubernetes Cluster begonnen, welches als Endpunkt für die automatisierte Datensammlung von Produktionsforschungsdaten genutzt werden kann. Das System beruht dabei ausschließlich auf produktionserprobter OpenSource Software um Hersteller Lock-Ins zu vermeiden und eine stabile Infrastruktur für das Forschungsdatenmanagement im Maschinenbau zu bilden. Dieser Vortrag geht auf die verwendete Hardware Infrastruktur, die verwendete Software und den allgemeinen Workflow des Systems ein.

Gelebtes Forschungsdatenmanagement in den Materialwissenschaften: Workflows zur Weiterentwicklung von Point-of-Care-Antigen-Tests

Patrick Altschuh (HS Karlsruhe)

Der Paradigmenwechsel in der computergestützten Materialwissenschaft hin zur datengetriebenen Forschung birgt datenintensive Arbeitsprozesse. Insbesondere die Datenkuration stellt hierbei im Forschungsalltag eine enorme Herausforderung dar und lässt sich nur durch eine flexible Forschungsdateninfrastruktur bewältigen. Am Beispiel von Point-of-Care-Antigen-Tests wird in diesem Vortrag die datengetriebene Charakterisierung und Weiterentwicklung von porösen Polymermembranen vorgestellt, wobei die datenintensiven Workflows mit der Forschungsdateninfrastruktur Kadi4Mat umgesetzt werden.

KaDI4Mat - Forschungsdateninfrastruktur für die Materialforschung

Michael Selzer (KIT)

Forschungsdatenmanagement (FDM) in den Ingenieurwissenschaften ist von zunehmender Bedeutung. Forschungsdaten in großem Umfang inklusiv entsprechender Metadaten und Ontologien abzuspeichern, auszutauschen und zugänglich zu machen stellt dabei eine Herausforderung dar. Daher entwickeln wir am Institut für Angewandte Materialien des KIT in enger Abstimmung mit dem NFDI4Ing-Archetyp Caden die Karlsruhe Data Infrastructure for Materials Science (Kadi4Mat) mit dem Ziel Forschungsdaten nach den FAIR-Prinzipien digital zugänglich zu machen. Dabei unterstützt die Datenaustausch- und Analyseplattform Kadi4Mat Kooperation zwischen Experimentatoren, Theoretikern und Simulanten vor allem in den Materialwissenschaften bei der Entwicklung neuartiger Materialien.

B4**Test Driven Development (TDD)**Sören Peters (TUBS)
Sven Marcus (TUBS)

Stetig komplexere Problemstellungen erfordern geeignete Methoden und Techniken zum Entwurf und der Implementierung von Software. Zudem unterliegt die Software im wissenschaftlichen Kontext einem kontinuierlichen Wandel. Ein Ansatz zum Umgang mit der steigenden Komplexität, bei gleichzeitig hoher Flexibilität, ist die testgetriebene Entwicklung (engl.: test driven development) - TDD. Die Besonderheit dabei: bei der testgetriebenen Entwicklung wird der Test jeweils vor der zu testenden Komponente geschrieben. In diesem Workshop sprechen wir über die Grundlagen von TDD und werden dann anhand eines Beispiels TDD gemeinsam praktizieren.

C1

Strukturierte Metadaten in den Verkehrswissenschaften

Matthias Fuchs (SLUB)
Susanne Arndt (TIB)
Maximilian Bäumler (TUD)
Stefanie Roski (SLUB)

Von Szenarien für das automatisierte Fahren bis hin zu Anpassung der Infrastruktur - belastbare empirische Daten aus Verkehrsbeobachtungen mit mobilen und stationären Kameras spielen für eine Reihe verkehrswissenschaftlicher Fragen eine wichtige Rolle. Da die Erhebung solcher Daten mit einem gewissen Aufwand verbunden ist, sollte ihre Nutzbarkeit sichergestellt werden. Ein gut strukturiertes Forschungsdatenmanagement kann dazu einen wichtigen Beitrag leisten. Hierzu gehört auch ein abgestimmtes Schema, mit dem die Objekte, Prozesse und Bedingungen nachvollziehbar beschrieben werden. Im Rahmen des Workshops wird ein Entwurf eines solchen Schemas und seine Nutzung für das Datenmanagement an der Technischen Universität Dresden vorgestellt. Dem sich darauf aufbauenden fachlichen Austausch zu Erhebungswerkzeugen und den damit einhergehenden Erfahrungen im Datenmanagement schließt sich ein Impuls und eine gemeinsame Diskussion zu den Herausforderungen und Möglichkeiten einer community-getriebenen Weiterentwicklung von Schemata zur Unterstützung der Forschungsarbeit an. Der Fokus wird auf die Diskussion mit den Community-Vertretern über die benötigten Rollen, Ressourcen und Prozesse einer solchen Entwicklung gelegt, um abschließend die nächsten Schritte im Kontext der NFDI4Ing festhalten zu können.

C2

Einführung in Git

Dennis Gläser (Univ. Stuttgart)

Today, Git is the most-widely used version control system for collaborative software development, in particular in the Open-Source community. Depending on the size and the release cycles of a project, developer teams may define different workflows on how new features should be incorporated and how releases should be managed. In this workshop, we want to discuss selected Git-Workflows and provide the Git basics that are required to realize them.

C3

Energiedaten für die studentische Ausbildung in python-Energiedatenanalyse gesucht

Johanna May (TH Koeln)

Energieeffizienzverbesserungen benötigen Energiedatenanalyse. Diese erlernen Studierende durch die Nutzung von jupyter notebooks. Dabei verwenden sie selbst gemessene Daten und offen verfügbare Datenquellen für elektrische Lastprofile. Der Beitrag zeigt auf, in welchen Bereichen die Arbeit mit offenen Daten schon gut gelingt. Außerdem nennt er Potenziale weiteren Austauschs, z. B. auch von industriellen Verbrauchsdaten aus öffentlich geförderten Projekten oder studentischen Messergebnissen aus Lehrprojekten andererseits.

**Forschungsdatenmanagement in einem
Großforschungsverbund: Erfahrungen aus dem
Exzellenzcluster POLiS**

Ephraim Schoof (KIT)

Im universitätsübergreifenden Exzellenzcluster POLiS (Post-Lithium-Speicherung) wird an Zukunftsbatterien geforscht, die leistungsfähiger, zuverlässiger und nachhaltiger als derzeitige Lithium-Ionen-Batterien sind. Der verlässliche Austausch von Forschungsdaten inklusive ihrer Metadaten spielt dabei eine zentrale Rolle für gewinnbringende Kooperationen zwischen Forschenden innerhalb des Clusters. In diesem Vortrag werden Herausforderungen, Lösungsansätze und Erfahrungen diskutiert, um ein nachhaltiges Forschungsdatenmanagement universitätsübergreifend zu etablieren und die Nachnutzung von Forschungsdaten zu stärken.

**Anforderungen für Forschungsdatenmanagement
und integriertes maschinelles Lernen zwischen
Medizin und Ingenieurwissenschaften**

Julian Grolig (KIT)

Interdisziplinäre Projekte, z.B. zwischen Medizin und Ingenieurwissenschaften, profitieren von Systemen für integriertes Forschungsdatenmanagement (FDM), Simulationsframeworks und Maschinelles Lernen (ML). Während solche Systeme im klinischen Umfeld selten sind, sind FDM-Plattformen, Simulationsframeworks und ML aktive Forschungsfelder in den Ingenieurwissenschaften. Aus dieser Ingenieursperspektive beschreiben wir die Anforderungen und Herausforderungen für FDM, die Integration von ML in der Medizin und wie ein zuverlässiges FDM-Tool wie Kadi4Mat für die Implementierung einer solchen Lösung verwendet werden kann.

C4

**Pace3D - Ein Framework für High Performance
Mikrostruktursimulation zur datengetriebene
Entwicklung neuer Materialien**

Britta Nestler (KIT)

Pace3D (Parallel Algorithms for Crystal Evolution in 3D) ist ein modulares Softwarepaket für großskalige parallele (3D+t) Simulationen von Phasenumwandlungsprozessen und Mikrostrukturausbildungen in mehrkomponentigen mehrphasigen und polykristallinen Materialsystemen unter Berücksichtigung multiphysikalischer Einflussgrößen wie Wärme- und Stoffdiffusion, Strömung, Thermo-/Chemo-Mechanik, Elektrochemie und Magnetismus. Weiterhin enthält das Softwarepaket Pace3D ein umfangreiches Kompendium an Pre- und Postprocessing Algorithmen sowie Visualisierungsmethoden mit Programmierschnittstellen zur Datenvorbereitung und Datenanalyse der dynamischen Mikrostrukturentwicklung. Die Data Science Methoden sind als Workflows kombinierbar und werden in der Forschungsdateninfrastruktur Kadi4Mat unter FAIR Prinzipien zur nachhaltigen Anwendung zur Verfügung gestellt. Das Framework wird unter der Verwendung der Versionsverwaltung GIT kontinuierlich weiterentwickelt und enthält diverse Möglichkeiten der Dokumentation für Entwickler und Anwender.

Bembel - The BEM-based engineering library

Sebastian Schöps (TUDA)
Felix Wolf (TUDA)
Jürgen Dölz
Michael Multerer

This talk discusses the challenges in developing and maintaining the research code Bembel. Bembel is the Boundary Element Method Based Engineering Library written in C++ to solve boundary value problems governed by the Laplace, Helmholtz or electric wave equation within an isogeometric framework, i.e., it uses spline basis functions. The software was developed as part of a cooperation between TU Darmstadt, the University of Basel and USI Lugano in several projects. It is currently supported by the DFG in the programme „Qualitätssicherung von Forschungssoftware durch ihre nachhaltige Nutzbarmachung“.

Semantik im Internet of Production: Erfahrungen aus 3 Jahren Exzellenzcluster

Johannes Theissen-Lipp (RWTH)
Iraklis Dimitriadis (FIT Fraunhofer)

Im Exzellenzcluster „Internet of Production“ forschen an der RWTH Aachen University über Hundert Doktorand:innen und mehrere Dutzend Professor:innen. Nach fast drei Jahren fassen wir Erfahrungen bei der Zusammenarbeit an semantisch adäquaten und kontextabhängigen Daten zusammen und regen zur Diskussion im ähnlich dimensionierten Projekt NFDI4Ing an. Neben Erfahrungsberichten aus der Ontology Expert Group werden insbesondere die Definition und praktische Anwendung von Best Practices für Bereiche wie Ontologien, Applikationsprofile und Interoperabilität kritisch beleuchtet. Einige Lessons Learned sowie ein Schlußschluss mit vergleichbaren Projekten rundet den Vortrag ab und lädt zu einer offenen Diskussion ein.

Sessions D**D1****Visualization of networks – analyzing and visualizing connections between (planned) NFDI consortia**

Évariste Demandt (RWTH)
Lukas Bossert (RWTH)
Sama Majidian (RWTH)

Currently, Germany's National Research Data Infrastructure spans a network of nine funded consortia from the first round and ten from the second round. This workshop enables you to visually display and analyze the network of consortia in your internet browser via a remote Jupyter Notebook. The workshop follows the tradition of literate programming. No prior experience in programming and no locally installed software needed - let's weave and tangle!

D2**Continuous Integration**

Dennis Gläser (Univ. Stuttgart)

Testing the functionality of a software project is a fundamental measure for quality assurance. However, the larger the test suite becomes, the more it imposes a burden on the developers that have to execute it. An automation of this process is therefore preferable, and in this workshop we want to outline how the GitLab CI/CD can be used to incorporate automated testing into a Git-Workflow to continuously guarantee the functionality of a software project.

Continuous Integration of Scientific Software

Tomislav Maric (TUDA)

The talk covers a straightforward workflow that uses version control, continuous integration, and data repositories to increase the reproducibility and findability of research data generated by scientific software. The workflow was developed for small academic research teams at universities without additional software engineering support and is therefore kept as simple as possible.

Making software citation easi(er) – The Citation File Format and its integrations

Stephan Druskat (DLR)

To credit research software work, enable access to software and reproducibility of research results, and adhere to Good Scientific Practice, software must be cited in publications. To make software citation possible, the respective metadata is provided by developers/RSEs, for example in the Citation File Format (CFF). Software citation based on CFF has recently become a lot easier due to format support on platforms such as GitHub, Zenodo and Zotero. In this talk I'll give a brief overview about the Citation File Format, recent developments, and new and upcoming platform integrations.

Erfahrungen mit dem Testen des Gewässergütemodells QSim

Jens Wyrwa (BAFG)

Die Softwarefamilie QSim, das deterministische Gewässergütemodell der Bundesanstalt für Gewässerkunde (BfG) wird vorgestellt. Die Herausforderung für das Testen dieser Software ergibt sich aus der großen Zahl von vielfältig gekoppelten Modulen. Möglichkeiten zur Entwicklung von Teststrategien werden diskutiert. Die unterschiedlichen Test-Anlässe werden präsentiert. Die Technik, einzelne Module, die sich nicht isolieren lassen, mit konsekutiven Tests zu überprüfen, wird vorgestellt. Um Fragen der Testbarkeit auch in der Diskussion weiter vertiefen zu können, werden drei praktisch aufgetretene Bugs näher erläutert.

D3

Data Stewards in Verbundforschung

Peter Pelz (TUDA)
Christian Schänzle (TUDA)
Nils Preuss (TUDA)

n.a.

Monitoring von Versuchsobjekten in einem Forschungsdatenmanagementsystem am Beispiel einer Tailored-Forming-Prozesskette

Iryna Mozgova (LUH)

Die Transparenz und Verfügbarkeit von Forschungsergebnissen ist ein wertvolles Element der wissenschaftlichen Arbeit im Ingenieurwesen. Ein wichtiger Aspekt der Arbeit in großen Verbundprojekten ist die Überwachung und Kontrolle des Status von Versuchsobjekten, insbesondere wenn mehrere Teilprojekte gemeinsam forschen, z. B. innerhalb einer technologischen Prozesskette.

In diesem Vortrag wird am Beispiel des Sonderforschungsbereichs 1153 "" Prozesskette zur Herstellung hybrider Hochleistungsbauteile durch Tailored Forming"", in dem ein interdisziplinäres Team neuartige Prozessketten zur Herstellung hybrider Bauteile erforscht, ein Ansatz zur Organisation des Forschungsdatenmanagements mit Hilfe eines Forschungsdaten- und Wissensmanagementsystems gemäß den FAIR-Datenprinzipien vorgestellt.

**Validation as a specific aspect of testing the
CORSIKA Air Shower Monte Carlo Simulation
package**

Ralf Ulrich (KIT)

It is very often the goal of research software to combine and connect various algorithms into more complex frameworks in order to tackle new scientific problems. While it is understood that testing and unit testing are crucial for the stability and long term maintenance of any modern software project, we propose to clearly distinguish to basic testing of the interface to an algorithm for fundamental code correctness, from the actual algorithm itself. Also because the latter often is a model-dependent and time-consuming endeavor. Such external algorithms essentially need to be "validated", which is not identical to be "unit tested", but this can also be done in a time-dependent and automatic way. We illustrate this at the example of the CORSIKA air shower simulation framework.

Sessions E

E1

**plotID, Toolkit zur Verknüpfung von Abbildungen
und Forschungsdaten**

Jan Lemmer (TUDA)
Martin Hock (TUDA)
Manuela Richter (TUDA)

Eines der wichtigsten Darstellungsmittel von wissenschaftlichen Daten sind Abbildungen und insbesondere Plots der (experimentellen) Daten. Die Nachvollziehbarkeit und Zuordnung der Daten in diesen Plots sind häufig schwierig, dies betrifft sowohl die Meta-Daten, die Roh-Daten als auch die Verarbeitungs-Skripte. Das Toolkit plotID unterstützt Forschende bei diesem Prozess. Abbildungen werden mit einer ID versehen und die zugehörigen Daten abgelegt. Im Rahmen des Vortrages wird das Konzept vorgestellt und der aktuelle Entwicklungsstand präsentiert.

**Datencontainer mit Provenienz: Einfacher
Datenaustausch mit verteilten Stakeholdern**

Frank Dressel (DLR)

In diesem Vortrag wird ein Datencontainer vorgestellt, welcher Provenienz und zugehörige Daten kombiniert. Dieser eignet sich besonders für a) Datenaustausch zwischen verteilten und unabhängigen Stakeholdern und b) Datenverwendung über einen langen Zeitraum hinweg. Für die Verwendung der Container werden keine spezialisierten Tools oder Services benötigt. Die Anforderungen/Einschränkungen sowie Pro und Kontra des gewählten Ansatzes werden diskutiert.

**Ein ontologiebasiertes Datenmanagementsystem
für die Produktionstechnik**

Iraklis Dimitriadis (FIT)
Martin Kretschmer (FIT)

Wir stellen eine Datenmanagement-Plattform vor, welche zurzeit im Fraunhofer Leitprojekt EVOLOPRO entwickelt wird. Ziel dieser Plattform ist es den prozessübergreifenden Informationsaustausch in der Produktionstechnik zu verbessern und die Implementierung von Digitalen Zwillingen zu unterstützen. Hierzu werden unterschiedliche Konzepte aus dem Semantic Web Bereich kombiniert und angewandt, um hochgeladene Datensätze mithilfe von Ontologien automatisch oder manuell zu annotieren und verknüpfen. Neben der web UI kann die Datenplattform auch über eine LDP spezifizierte REST API bedient werden und ermöglicht somit auch einen automatisierten Datenaustausch.

E2**Testing open source Python-centric research software**

Hans Fangohr (MPSD)

This presentation will contextualise automatic tests within software engineering for research software, explain underlying ideas with small educational examples and discuss the use of these ideas in open source software projects. While the focus is on software written in Python, the underlying concepts are transferable to other languages.

Continuous testing in HPC fortran: lessons learned from TROVEJamie Quinn (UCL)
Anastasias Georgoulas (UCL)

TROVE, a molecular simulator, is a legacy, HPC code written in non-standard fortran, which presented a unique challenge when we wished to implement tests prior to major upgrades. In this talk we will share our experiences using the unit testing framework pFUnit, developing our own end-to-end regression testing framework, and setting up automatic testing through the continuous integration provider Github Actions. This combination of different tests run automatically allows us greater confidence in making significant changes to the codebase.

Continuous struggle: precise, performant and portable code development in HPC

Ivo Kabadshow (FZ Juelich)

Writing code is easy, developing software is hard. This is especially true in HPC, since the software does not only need to run on a single system for which it can be tuned, but must run efficiently and correct on a multitude of hardware architectures. Therefore, continuous testing becomes mandatory for multi-user code bases. In this presentation we will take a closer look at continuous testing for HPC libraries and the combinatorial explosion of tests due to the diverse HPC ecosystem.

E3**Was Forschende wollen - Erfahrungsbericht aus FDM-Beratungen**

Andreas von der Dunk (SLUB)

Die FDM-Beratungen der Kontaktstelle Forschungsdaten in Dresden orientieren sich streng an den Bedürfnissen der Forschenden und basieren auf einem eigens dafür entwickelten Beratungskonzept. Der Vortrag synthetisiert die Erkenntnisse aus über 100 Beratungsmandaten: welche Anforderungen haben die Forschenden an das FDM und welche Lösungskonzepte werden angewendet.

Akzeptanz- und Motivationsförderung mit FDM-User StoriesAmelie Metzmacher (RWTH)
Tobias Hamann (RWTH)
Mario Moser (RWTH)

Forschungsdatenmanagement scheitert in der Praxis häufig an der mangelnden Akzeptanz und geringen Motivation der Forschenden, da diese oft nur den erhöhten Zeitaufwand wahrnehmen. Bislang gibt es wenige anschauliche Beispiele, die aufzeigen, welche Vorteile ein adäquates Forschungsdatenmanagement implizieren. Im Workshop werden gemeinsam mit den Teilnehmenden Best und Worst Practices strukturiert gesammelt und im Datenlebenszyklus eingeordnet. Die Beispiele werden anschließend zu User Stories entwickelt, in denen gezielt das "Warum" und das "Wie" eines guten Forschungsdatenmanagements adressiert werden.



Ausrichter der NFDI4ing Konferenz 2021:

TU Braunschweig: Manfred Krafczyk, Jan Linxweiler

Mit Unterstützung von:

SLUB Dresden: Stefanie Roski, Matthias Fuchs

TU Dresden: Regine Gerike, Marco Berger

Wir bedanken uns bei allen Autoren und Sprechern für die eingereichten Beiträge. Für die Inhalte der sind ausschließlich die jeweiligen Autoren bzw. Sprecher verantwortlich.

Weitere Informationen zur Nationalen Forschungsdateninfrastruktur für die Ingenieurwissenschaften NFDI4ing unter:

www.nfdi4ing.de

Weitere Informationen zur Nationalen Forschungsdateninfrastruktur und weiteren Konsortien unter:

<https://www.dfg.de/foerderung/programme/nfdi/>

Versionsstand des Dokumentes: 09.11.2021 10:03:27