
Nationale Forschungsdateninfrastruktur für die Ingenieurwissenschaften (NFDI4Ing)

Britta Nestler^{1c}, Peter F. Pelz^{2c}, Robert H. Schmitt^{3c}, Marco Berger^{4a}, Hauke Dierend^{5a}, Benjamin Farnbacher^{6a}, Bernd Flemisch^{7c}, Dennis Gläser^{7a}, Ina Heine^{3c}, Nils Hoppe^{6a}, Gerald Jagusch^{2a}, Roland Lachmayer^{5c}, Jan Lemmer^{2a}, Jan Linxweiler^{8a}, Amelie I. Metzmacher^{3a}, Iryna Mozgova^{5c}, Nils Preuß^{2a}, Manuela Richter^{2a}, Stefanie Roski^{9a}, Hartmut Schlenz^{10a}, Michael Selzer^{1a} und Christian Stemmer^{6c}

^aAutor:innen

^cKontribuierende

¹Karlsruher Institut für Technologie

²Technische Universität Darmstadt

³Rheinisch-Westfälische Technische Hochschule Aachen

⁴Technische Universität Dresden

⁵Leibniz Universität Hannover

⁶Technische Universität München

⁷Universität Stuttgart

⁸Technische Universität Braunschweig

⁹Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden

¹⁰Forschungszentrum Jülich

NFDI4Ing ist ein 2017 gegründetes Konsortium mit dem Ziel, Wissenschaftler:innen aller Disziplinen zu ermöglichen, ingenieurwissenschaftliche Forschungsprozesse in ihrer Gesamtheit nachvollziehen oder reproduzieren zu können. Die Besonderheit an NFDI4Ing ist der Aufbau, welcher sich in drei Bereiche aufteilt. Die Archetypen, die an den methodischen Bedarfen ausgerichtet sind, die Community Cluster und die Base Services. NFDI4Ing erarbeitet technologische Methoden und Lösungen, bietet Aus- und Weiterbildungsprogramme und trägt zur Verbreitung des Forschungsdatenmanagements (FDM) in den Ingenieurwissenschaften bei.

1 Einleitung

NFDI4Ing hat das Ziel, Forschungsdaten in den Ingenieurwissenschaften FAIR zu machen – auffindbar, zugänglich, interoperabel und nachnutzbar [1]. Dabei stellt sich das Konsortium drei zentralen Herausforderungen [2]: (i) Bildung zur Sicherstellung der Datenkompetenz von Anfang an, (ii) Validierung von technologischen Lösungen und Methoden für Forschungsdaten sowie (iii) Erprobung von Konzepten, die Förderung von Data Governance und der Datenkuration. Die meisten TU9-Universitäten sowie das Deutsche Zentrum für Luft- und Raumfahrt (DLR), das Forschungszentrum Jülich, als auch

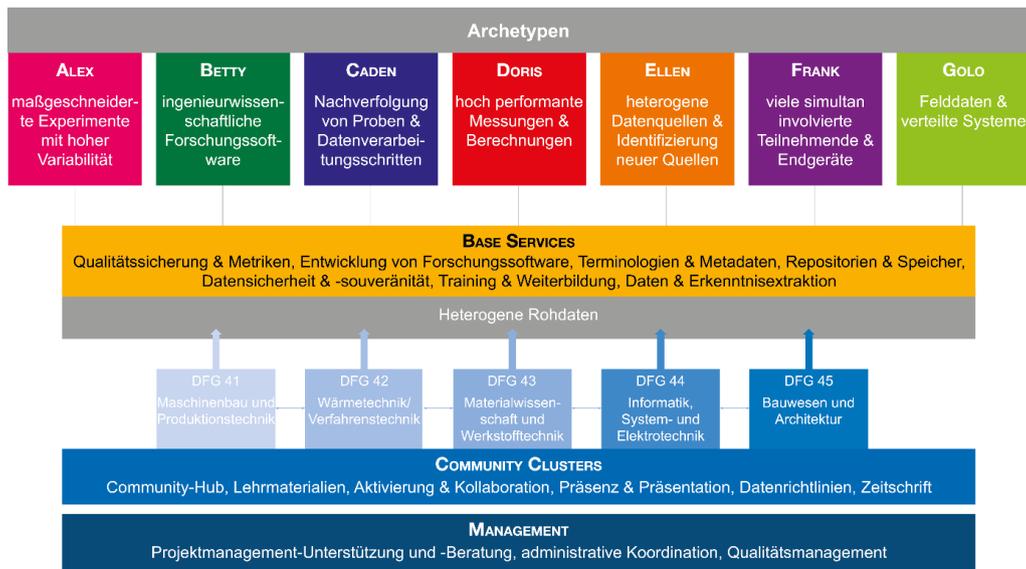


Abbildung 1: Aufbau von NFDI4Ing.

das Leibniz-Informationszentrum Technik und Naturwissenschaften Hannover (TIB) sind Gründungsmitglieder des Konsortiums.

Der Aufbau von NFDI4Ing orientiert sich an dem Aufbau eines Unternehmens mit Entwicklung, Produktion, Vertrieb und der Unternehmensleitung, s. Abb. 1. Die Entwicklungsabteilung wird durch die einzelnen Archetypen repräsentiert. In diesem modularen und methodenorientierten Ansatz bearbeitet jeder Archetyp spezifische Aufgaben und Ziele. Für den Vertrieb sind die Community Cluster (CC) zuständig. Sie bilden die Schnittstellen zu den jeweiligen Fachgebieten in den Ingenieurwissenschaften (DFG Fachgebiete 41-45). Die Base Services stehen für die Produktion, indem sie zentrale Dienste für die Archetypen und die Community Cluster zur Verfügung stellen. Weiterhin werden Dienste gepflegt und entwickelt, die von den Community Clustern als relevant erachtet werden. Das Management koordiniert und verwaltet die anderen Organisationsstrukturen.

2 Organisationsstrukturen

Archetypen

Die große Vielfalt ingenieurwissenschaftlicher Problemstellungen führt zu hochspezialisierten, individualisierten Forschungsansätzen und -methoden. Gleichzeitig lassen sich interdisziplinäre Gemeinsamkeiten auf der Ebene der verwendeten Forschungsmethoden und -prozesse erkennen [3]. Die Archetypen der NFDI4Ing, von ALEX (maßgeschneiderte Experimente) bis GOLO (Felddaten), harmonisieren diese Gemeinsamkeiten und repräsentieren die Vielfalt der FDM-Anforderungen im Ingenieurwesen.

1. ALEX

Der Archetyp ALEX repräsentiert Forschende, die physico-chemische Effekte technischer Systeme mit Hilfe maßgeschneiderter Experimente untersuchen. Diese können sowohl real als auch software-gestützt sein. In ihrer täglichen Arbeit konkurrieren Kompatibilität und Flexibilität, so werden häufig einzigartige Teillösungen für den jeweiligen Versuch entwickelt. Daher sind modulare, selbstdokumentierende und wiederverwendbare Softwarebausteine notwendig. Ziel ist, Konzepte und Best-Practice Beispiele für modulare Teillösungen zu erarbeiten. Dadurch werden die Nachvollziehbarkeit der Systemfunktionen, des Datenflusses und der Konfiguration verbessert. Weitere Schwerpunkte sind die Selbstdokumentation der Software und die automatische Generierung von maschinenlesbaren Metadaten. Dies soll zukünftig sowohl die Veröffentlichung und Wiederverwendung von Forschungsdaten fördern als auch die Entkopplung von Code und Daten erleichtern.

2. BETTY

Der Archetyp BETTY repräsentiert Entwickelnde von Forschungssoftware, bspw. für die Simulation physikalischer Systeme mit eigens entwickelten Methoden. Um die Transparenz der Forschung und die Reproduzierbarkeit der Ergebnisse zu gewährleisten, muss die Forschungssoftware nachnutzbar publiziert werden. Ziel der Task Area (TA) ist es deshalb, Forschende mit Werkzeugen und Wissen für die Entwicklung validierter und qualitätsgesicherter Ingenieurs-Forschungssoftware auszustatten. Dies beinhaltet Guidelines und Standards für die Entwicklung und Veröffentlichung von Forschungssoftware, sowie die Definition und/oder Anwendung von Metadatenstandards zur Beschreibung der implementierten Methoden und der Abhängigkeiten von anderen Softwarepaketen. Die in der TA entwickelten Ansätze werden zunächst in drei Pilotanwendungen aus den Ingenieurwissenschaften exemplarisch umgesetzt und durch Hinzunahme neuer Anwendungen systematisch erweitert und verallgemeinert.

3. CADEN

Ein zentrales Interesse des Archetyps CADEN ist das sog. Provenance-Tracking von Proben und Daten. Die zentrale Anforderung besteht darin, Datenentitäten (d. h. sowohl Daten als auch Metadaten) zu speichern und Parameter von Aktivitäten (z. B. Temperaturen, Drücke, Simulationsparameter) strukturiert und nachverfolgbar zu erfassen. Darüber hinaus müssen die Verknüpfungen der Entitäten zur Beschreibung einer Graphentopologie erstellt werden. Der Graph kann bei einer Vielzahl von Prozessschritten sehr komplex und nicht-linear sein (d. h. Verzweigungen und Abzweigungen enthalten). Eine weitere Herausforderung in diesem Archetyp ist die Kooperation unterschiedlicher Einrichtungen. Institutionen speichern Metadaten bisher isoliert voneinander und eine maschinenverarbeitbare Verknüpfung zwischen den Fragmenten des Graphen ist bisher nicht etabliert. Eine Möglichkeit dies zu lösen, ist der Einsatz einer einheitlichen Forschungsdateninfrastruktur, wie z. B. *Kadi4Mat* [4] oder *eLabFTW* [5].

4. DORIS

Im Archetyp DORIS werden Forschende repräsentiert, die hochgenaue und hochauflösende Messungen durchführen, und/oder Simulationen mit Hilfe von Höchstleistungsrechnern (HPMC) ausführen. Durch die Größe der erzeugten Daten (hunderte TB bis PB) sind diese nicht mobil, sondern nur über lokale und personalisierte Accounts bei den Re-

chenzentren abrufbar und nicht in Datenbanken oder Repositorien aufzufinden. Durch die Entwicklung einer interoperablen HPMC-Ontologie sowie passgenauer Softwarelösungen für die Auffindbarkeit und den Zugang zu HPMC-Daten und -Metadaten, soll der Zugriff durch Drittnutzende ermöglicht werden. Um diese Werkzeuge zu etablieren und deren Nutzung zu fördern, werden Forschende mit Workshops und Best-Practice-Guides unterstützt. Durch breiten Zugang zu interoperablen HPMC-Daten werden neue Forschungsansätze ermöglicht (Forschung an Daten), dazu zählen die Bereiche maschinelles Lernen, künstliche Intelligenz und neuronale Netzwerke.

5. ELLEN

Der Fokus des Archetypen ELLEN liegt auf komplexen Systemen, welche durch eine große Anzahl von multidisziplinären Abhängigkeiten charakterisiert werden. Diese sind dem Bereich der Datenwissenschaften zugeordnet, bspw. modellbasierte Simulationen und Optimierungsrechnungen. Die Berechnungen sind typischerweise sehr datenintensiv und erfordern Informationen aus vielen verschiedenen Disziplinen. Das Ziel der TA ist es, Ingenieur:innen bei dem Such- und Implementierungsprozess zu unterstützen, die Anzahl potenzieller Datenquellen zu erhöhen, deren Integrationsgrad zu steigern und den Zeitaufwand für den Suchprozess zu reduzieren. Dazu sollen methodische Konzepte und deren Softwareimplementierungen zur Verfügung gestellt werden, welche ebenfalls in der Lage sind, fehlende Daten zu generieren.

6. FRANK

Der Archetyp FRANK repräsentiert Forschende, die sich interdisziplinärer Forschungsmethoden, z. B. aus der Produktionstechnik, der Ergonomie oder dem Wirtschaftsingenieurwesen, bedienen. Diese resultieren in einer Vielzahl verschiedener und heterogener, z. T. zu verknüpfender, Datenquellen, z. B. Fertigungsdaten in Echtzeit, Simulations- oder Probandendaten. Eine große Herausforderung bezieht sich somit auf die Synchronisation und das Zugriffsmanagement dieser heterogenen Daten in interdisziplinären Forschungsteams. Die Aufgaben der TA bestehen in der Etablierung gemeinsamer Ontologien und der Neugestaltung eines entscheidungsunterstützenden FDM-Frameworks, das u. a. die Datenrückverfolgbarkeit optimiert. Diese tragen zu der übergeordneten Zielsetzung bei, die Akzeptanz von FDM durch verbesserte Anwendbarkeit zu erhöhen und somit interdisziplinäre Forschungskollaborationen zu fördern.

7. GOLO

Der Archetyp GOLO repräsentiert Forschende, die sich mit der Planung, Erfassung, Klassifizierung und Analyse von Felddaten technischer Systeme beschäftigen. Felddaten dienen dazu, Modelle eines technischen Systems zu optimieren sowie detaillierte Analysen unter realen Betriebsbedingungen entlang des Lebenszyklus eines technischen Systems durchzuführen. Im Fokus steht die Entwicklung von Methoden und Werkzeugen zur Erfassung und Wiederverwendung der Felddaten. Anhand des Konzeptes eines Digitalen Zwillings erfolgt die Datenrepräsentation und -strukturierung nach den FAIR-Data-Prinzipien [6]. Zudem werden Checklisten und Best-Practice-Ansätze für die Arbeit mit Felddaten definiert. Durch eine Standardisierung von Schnittstellen und Abläufen soll die Wiederverwendung von Felddaten gefördert und erleichtert werden.

Base Services

Es sind sieben Base Services identifiziert worden, die für alle Archetypen und damit für die übergreifenden Leitziele von NFDI4Ing relevant sind und zentral entwickelt und erbracht werden: (i) Bereitstellung von Datenqualitätssicherungsprozessen, den entsprechenden Werkzeugen und Datenqualitätsmetriken, um Daten FAIR [6] zu machen; (ii) Unterstützung bei der Entwicklung von Forschungssoftware; (iii) Bereitstellung einfach zu bedienender und verständlicher Metadatenwerkzeuge und die Etablierung detaillierter ing.-wiss. Terminologien; (iv) sichere Speicherung und Langzeitarchivierung von Daten sowie die Möglichkeit, diese zu teilen oder zu publizieren; (v) Authentifizierungs-, Autorisierungs- und Rollenmanagementinfrastruktur (notwendig für Daten aus industrienahen Forschungsprojekten); (vi) Konzepte und Materialien für die Ausbildung; (vii) Techniken der Datenextraktion und Wissensentdeckung in der technischen Literatur.

Community Cluster

Der Aufgabenbereich Community Cluster ist das Vertriebsorgan von NFDI4Ing und verfolgt zwei Aufgaben. Erstens wird in sechs Arbeitsbereichen an Maßnahmen zur Verbreitung der entwickelten Dienstleistungen und Best-Practices gearbeitet. Diese sind: (i) Kommunikation, (ii) Lehre, (iii) Zusammenarbeit, (iv) Partizipation, (v) Standardisierung und (vi) Journal. Zweitens dient er als Schnittstelle zu den fünf DFG-Forschungsgebieten 41-45 und deren Bedürfnissen, wobei jedes Gebiet durch Co-Spokespersons repräsentiert wird. Die Forschungsgebiete werden im Community Cluster gezielt eingebunden, um die entwickelten Lösungen umzusetzen:

41 Maschinenbau und Produktionstechnik

Die Community 41 repräsentiert Fachdisziplinen, z. B. Mechanik, Textiltechnik, Produktionsorganisation und Betriebswissenschaften, die über die Verbindung von Grundlagen- und angewandter Forschung industrielle Herausforderungen adressieren. Die Community zeichnet sich durch zahlreiche Querschnittsthemen aus, die z. B. in Verbänden und Vereinen wie der Wiss. Gesellschaft für Produktionstechnik (WGP) gebündelt werden. Diese Vereinigungen werden im Community Cluster gezielt eingebunden, um einerseits die Bedürfnisse der Community zu ermitteln und andererseits die in NFDI4Ing entwickelten Lösungen in die Anwendung zu bringen.

42 Thermofluidmechanik und Verfahrenstechnik

Die Einbindung der Community erfolgt über Projekte, Vorträge und Workshops. Neben der bestehenden Anbindung an Verbundforschung werden Transferprojekte mit der Industrie durchgeführt. Die Zusammenführung und der Austausch mit den Teilcommunities und deren fachliche Perspektiven findet z. B. auf der PAAT 2020 [6], der 2. jährlichen NFDI4Ing Konferenz [7] mit dem Schwerpunkt FDM in der Thermofluidik, statt. In 2021 wird ein Community Board eingerichtet, welches hilft die Lösungen von NFDI4Ing in die Community 42 hinein zu tragen.

43 Materialwissenschaft und Werkstofftechnik

In den Materialwissenschaften ist die Umsetzung digitaler und automatisierter Arbeitsflüsse für eine beschleunigte und fortschrittliche Entwicklung neuer Materialien von großer Bedeutung. Repräsentative Arbeitsflüsse verknüpfen etablierte Algorithmen der Datenvorbereitung und -analyse. Die datengetriebenen Erkenntnisse und Informationen können als Graphen veranschaulicht werden und ermöglichen eine gezielte, anwendungsorientierte, maßgeschneiderte Auslegung neuer Werkstoffe.

44 Informatik, System- und Elektrotechnik

Die Community umfasst Disziplinen aus den Fachrichtungen Informatik, System- und Elektrotechnik. Fachübergreifende Querschnittsfragen ergeben sich dabei bspw. bei Anforderungen zur Handhabung großer Datenmengen, dauerhafter Datenaufbewahrung und -bereitstellung sowie disziplinübergreifender Modellentwicklung. Adressiert werden diese und weitere Fragestellungen durch Vorträge, Workshops und weitere partizipative Veranstaltungsformate.

45 Bauwesen und Architektur

Die Aktivierung der Community erfolgt u. a. durch Einbeziehen des Fakultätentags für Bauingenieurwesen, Geodäsie und Umweltingenieurwesen (FTBGU) sowie der Dekane- und Abteilungsleiterkonferenz für Architektur, Raumplanung und Landschaftsarchitektur (DARL). Dabei liegt ein besonderer Fokus auf dem Austausch der Communities mit den Archetypen, der bspw. durch regelmäßige Umfragen und das Einrichten von Community Boards realisiert wird.

Special Interest Groups (SIGs)

SIGs behandeln NFDI4Ing übergreifenden Themen bzw. Interessen und stellen die Querkommunikation sicher. Diese erfolgt sowohl zwischen den Organisationsstrukturen innerhalb von NFDI4Ing (Archetypen, Base Services, Community Cluster) als auch über die Grenzen von NFDI4Ing hinaus, bspw. durch die Einbindung externer Expert:innen. Bisher wurden zwei SIGs gegründet: (i) Metadaten und Ontologien und (ii) Basic FDM Training & Weiterbildung.

3 Ausblick

Wichtige Ergebnisse der Arbeiten von NFDI4Ing werden in einem neu geschaffenen Journal für Forschungsdatenmanagement im Ingenieurwesen sowie auf der Webseite des Konsortiums (nfdi4ing.de) veröffentlicht und auf der jährlichen NFDI4Ing-Konferenz vorgestellt.

Danksagung

Die Autor:innen und die Kontribuierenden möchten sich bei Bund, Ländern und bei der Gemeinsamen Wissenschaftskonferenz (GWK) für die Förderung und Unterstützung im Rahmen des Konsortiums NFDI4Ing bedanken. Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 442146713.

Literaturverzeichnis

- [1] Mark D. Wilkinson u. a. The FAIR Guiding Principles for scientific data management and stewardship. 3:160018, 2016. Journal Article. doi:<https://doi.org/10.1038/sdata.2016.18.eprint:26978244>.
- [2] Robert H. Schmitt u. a. 209 .YHAu6j9CSUk. NFDI4Ing - the National Research Data Infrastructure for Engineering Sciences, 2020. <https://zenodo.org/record/4015201#.YHAu6j9CSUk>.
- [3] Gerald Wolfgang Jagusch and Nils Preuß. Umfragedaten zu "NFDI4Ing - Rückmeldung aus den Forschungscommunities", 2019. doi:<https://doi.org/10.25534/TUDATALIB-104>.
- [4] Nico Brandt, Lars Griem, Christoph Herrmann, Ephraim Schoof, Giovanna Tosato, Yinghan Zhao, Philipp Zschumme, and Michael Selzer. Kadi4Mat: A Research Data Infrastructure for Materials Science. 20, 2021. doi:<https://doi.org/10.5334/dsj-2021-008>.
- [5] Nicolas CARPi, Alexander Minges, and Matthieu Piel. eLabFTW: An open source laboratory notebook for research labs. 2(12):146, 2017. doi:<https://doi.org/10.21105/joss.00146>
- [6] DECHEMA e.V. Jahrestreffen der ProcessNet-Fachgemeinschaft Prozess-, Apparate- und Anlagentechnik (PAAT), 2020. https://dechema.de/en/PAAT2020_Programm/_/AK_Progr_2020.pdf(besucht am 09. 04. 2021).
- [7] TU Darmstadt. Gepflegt, geteilt, digital lesbar, 2020-12-08.https://www.tu-darmstadt.de/universitaet/aktuelles_meldungen/einzelansicht_287808.de.jsp (besucht am 09. 04. 2021).