

# Nationale Forschungsdateninfrastruktur für und mit Computer Science (NFDIxCS)

<https://nfdixcs.org> / Twitter: @nfdixcs

*Sprecher: Prof. Dr. Michael Goedicke,*

*Uni Duisburg-Essen und Gesellschaft für Informatik e.V., [mg@uni-due.de](mailto:mg@uni-due.de)*

*Sprecherin: Prof. Dr. Ulrike Lucke,*

*Uni Potsdam und Gesellschaft für Informatik e.V., [ulucke@uni-potsdam.de](mailto:ulucke@uni-potsdam.de)*

## Ziele

1. Das Konsortium NFDIxCS möchte (a) die Umsetzung der FAIR Data Principles für Informatik-Forschungsdaten sowie Software-Artefakte fördern, (b) die Zitierbarkeit von Software und Informatik-Daten vereinfachen und damit (c) die Publikationsprozesse und -kultur sowohl in der Informatik als auch in ihren Anwendungen modernisieren. NFDIxCS bietet ein Forum für die Diskussion über Formate von Informatik-Forschungsdaten, Metadatenformaten und Semantiken. Es wirkt auf allgemein akzeptierte Standards, insbesondere für die nachhaltige Speicherung, das Auffinden und Zurverfügungstellen von Informatik-Forschungsdaten hin.

2. Die Informatik ist dabei eine wichtige Schnittstelle in der Wissenschaft: Zum einen ist sie eine eigene Disziplin, die zunehmend große Datenmengen erzeugt und für die eigene Forschung benötigt. Dies gilt nicht nur für Bereiche wie Big Data oder Künstliche Intelligenz / Maschinelles Lernen, sondern auch für den Betrieb von HPC-Systemen, Rechnerarchitektur etc. Zum anderen spielt die Informatik eine wichtige Rolle in anderen wissenschaftlichen Disziplinen. Zahlreiche Verfahren und Prozesse, die neue Erkenntnisse liefern, wären, ohne die Informatik nicht zu realisieren. Auch hier spielen große und größte Datenmengen eine Rolle, welche wiederum eine großartige Möglichkeit bieten, auch genuine Informatikmethoden weiter zu entwickeln.

3. NFDIxCS möchte die Erfahrungen und das Wissen der Informatik-Community zu Systemarchitekturen, Prozessen, Standards für Interoperabilität, datenorientierte wissenschaftliche Publikations- und Kommunikationssysteme mit allen interessierten Wissenschaftsbereichen bzw. allen in der NFDI vertretenden Konsortien teilen. Diese bidirektionale Zielstellung wird durch das „x“ im Akronym repräsentiert. In diesem Sinne möchte das Konsortium auch weitere Partner gewinnen und auf Augenhöhe kooperieren, um die oben skizzierten Ziele zu erreichen und die zugehörigen Dienste nachhaltig zu betreiben.

## Besondere Anforderungen der Informatik an FDM

Informatik-Forschung ist in den beiden oben genannten Rollen experimentell, beobachtend, analytisch, interpretierend, gestaltend und modellbildend tätig. Dementsprechend vielfältig sind die Forschungsdaten, die in der Informatik entstehen. Es existieren sowohl stark strukturierte (Beispiel: Messdaten in tabellarischer Form), semi-strukturierte (Beispiel: Logdateien aus dem Betrieb von verteilten Rechnersystemen, Netzwerken und IT-Security aber auch Arten von Aussagen in und über Software) als auch unstrukturierte Forschungsdaten (Beispiel: annotierte Bild- oder Textdatenbanken als Basis für Methodenentwicklung im Machine Learning, Daten aus Untersuchungen zu HCI und E-Learning aber auch technische und theoretische Informatik-Daten). Die fachspezifischen Prozesse zur Erzeugung, Verwaltung, Verarbeitung und Veröffentlichung dieser Daten reflektieren diese Vielfalt.

Eine Ende 2019 durchgeführte Umfrage unter den 14 Fachbereichen der Gesellschaft für Informatik e.V. (GI) zeigt, dass die von Informatikerinnen und Informatikern produzierten Forschungsdaten neben ihrer Struktur auch in Umfang stark variieren.

Die hieraus resultierenden besonderen Herausforderungen für NFDIxCS liegen darin, dass neben den eigentlichen (Roh-)Daten auch die damit verbundenen Softwareartefakte zur Erzeugung, Vorverarbeitung, Analyse und Darstellung der Daten sowie die Umgebungs- und Ausführungsinformation der jeweiligen Experimente mit abgespeichert werden müssen. Dies muss in einer Form geschehen, in der auch nach vielen Jahren und neuen Versionen verwendeter Betriebssysteme, Programmierumgebungen, Datenbanken etc. die Daten wiederhergestellt und – wenn irgend möglich – die in Frage stehenden Experimente wiederholt werden können. Damit einher geht, dass vor allem für die Validität und Qualität von Publikationsprozessen ein „Einfrieren“ der Daten und der dazugehörigen Umgebungs- und Ausführungsinformation gewährleistet sein muss. In der Regel ist es auch notwendig, die verschiedenen Entwicklungsinformationen bei dieser Speicherform mit zu berücksichtigen. Unter Einbeziehung der Anforderungen aus Heterogenität in Mengengerüst und Struktur und der Forderung nach Einbeziehung lokaler Lösungen entsteht das Bild eines (sehr) komplexen verteilten Informationssystems, das aus lose gekoppelten erweiterbaren Teilarchitekturen der Daten- und Infrastruktur-Anbieter bestehen wird.

Es besteht daher der Bedarf an einem auf Informatik-Aspekte fokussierten NFDI-Konsortium, da die bisherigen Konsortien die spezifischen Anforderungen, welche Forschungsdaten aus der Informatik mit sich bringen, nicht ausreichend berücksichtigen.

## Geplante Maßnahmen und Services

### **Community Building, Kompetenzaufbau & Dialog**

Die Entwicklung praxistauglicher Standards und Prozeduren muss praxisnah durch involvierte Informatikerinnen und Informatiker erfolgen. NFDIxCS versteht sich vor diesem Hintergrund auch als Debatten-Katalysator und Ermöglicher von Diskussion und Dialog innerhalb der Informatik-Forschungscommunity – auch jenseits der Gesellschaft für Informatik. Mit ihren rund 20.000 persönlichen und 250 korporativen Mitgliedern bietet die Gesellschaft für Informatik e.V. (GI) als größte und wichtigste Fachgesellschaft für Informatik im deutschsprachigen Raum bestehende Strukturen, um Diskussionen zu bündeln. Hier existieren bereits seit längerer Zeit Gremien, die sich mit FDM beschäftigen. In den 14 Fachbereichen und zahlreichen Fachgruppen der GI können Besonderheiten der einzelnen Disziplinen in einem Bottom-Up-Prozess diskutiert und anschließend in einem fachbereichsübergreifenden NFDI-Arbeitskreis (AK NFDI) aggregiert werden. Außerdem wird das Thema Kompetenzaufbau in Form von Aus- und Weiterbildung adressiert. Weitere informatik-nahe Fachgesellschaften und Vereinigungen und Partner sollen auf Augenhöhe eingebunden werden u.a. de-RSE<sup>1</sup>.

### **Standardentwicklung**

NFDIxCS möchte auf die Entwicklung von allgemein akzeptierten Standards für die Speicherung von Metadaten, Daten-Austausch, Speicherung und Archivierung hinwirken und hierfür ein geeignetes Forum bieten. Hierzu gehören die Einbindung der Fach- und Arbeitsgruppen zu Metadaten, Datenstrukturen, Softwarearchivierung und moderner Publikationspraxis sowie disziplinübergreifende Standardisierungsprozesse im Dialog mit anderen NFDI-Konsortien.

---

<sup>1</sup> de RSE: Research Software Engineers <https://de-rse.org/de/>

Auch soll der Dialog mit fachspezifischen Verlagen, Journalen und Konferenzen über die Publikationspraxis von Daten und Software bzw. Ergänzungen der Publikationsverfahren um Reviews für Daten und Software und die persistente Referenzierung von Daten und Software geführt werden. Die GI ist ein langjährig erfahrener Träger für solche Aushandlungsprozesse.

## **Infrastruktur- und Service-Provider**

Um den Aufbau einer offenen Informatik-Forschungsinfrastruktur (Open Access, Open Science, Open Data) zu beschleunigen, sollen die Rechenzentren der GDWG, KIT und Leibniz-rechenzentrum Infrastruktur für die systematische Speicherung von Forschungsdaten und entsprechender Metadaten zur besseren Verarbeitung, Auffindbarkeit, Nachvollziehbarkeit und Reproduzierbarkeit unter den FAIR Data Principles bereitstellen. Eine Partnerschaft mit dem DFN ist besprochen, um Identitäten und ein adäquates Rechte- und Rollenkonzept zur Verfügung zu stellen. Auch soll Software und ihre jeweiligen Entwicklungskontexte und Ausführungsumgebungen (inkl. Software-Bibliotheken, Umgebung z. B. verwendete Hardware) archiviert werden. So sollen Softwareartefakte langfristig zur Verfügung stehen.

## **Partner und internationale Anbindung**

Die Mitglieder und Partner des Konsortiums haben einen großen Erfahrungsschatz in der Definition, Implementierung und Betrieb komplexer Service-orientierter Systeme.

Die Umsetzung soll in Abstimmung mit der europäischen und internationalen Informatik-Forschungsgemeinschaft geschehen. Auch die Software Heritage Foundation, das Alan Turing Institute und weitere internationale Informatik-Fachgesellschaften sollen als relevante Partner eingebunden werden.

## **Querschnittsthemen und Schnittstellen**

NFDIxCS hat sechs relevante Querschnittsthemen identifiziert, zu denen die Informatik wichtige Beiträge liefern kann: Security & Privacy; Usability; Distribution, Liason & Connectivity and Interoperability; Persistent Storage & Scalability; Evolution. Wir sind Mitzeichner der Leipzig-Berlin-Erklärung<sup>2</sup>.

---

<sup>2</sup> Maik Bierwirth, Frank Oliver Glöckner, Christian Grimm, Sonja Schimmler, Franziska Boehm, Christian Busse, ... Heike Neuroth. (2020, June 15). Leipzig-Berlin-Erklärung zu NFDI-Querschnittsthemen der Infrastrukturentwicklung. Zenodo. <http://doi.org/10.5281/zenodo.3895209>

**Vorgesehene Mitglieder des Konsortiums (Co-Sprecherinnen/Co-Sprecher und die weiteren, beteiligten Institutionen):**

<b>Co-Sprecher/in</b>	<b>Zugehörige Institution</b>
Prof. Dr. Michael Goedicke Professur Specification of Software Systems (S3) <a href="mailto:mg@uni-due.de">mg@uni-due.de</a>	paluno – The Ruhr Institute for Software Technology Universität Duisburg-Essen Gerlingstraße 16, 45127 Essen
Prof. Dr. Ulrike Lucke Professur Komplexe Multimediale Anwendungsarchitekturen <a href="mailto:ulrike.lucke@uni-potsdam.de">ulrike.lucke@uni-potsdam.de</a>	Universität Potsdam Institut für Informatik & Computational Science August-Bebel-Straße 89, 14482 Potsdam
Prof. Dr. Hannes Federrath Professur für Sicherheit in verteilten Systemen / Präsident der GI <a href="mailto:federrath@informatik.uni-hamburg.de">federrath@informatik.uni-hamburg.de</a>	Uni Hamburg / Gesellschaft für Informatik e.V. Ahrstraße 45, 53175 Bonn
Dr. Frank Löffler (angefragt) <a href="mailto:frank.loeffler@uni-jena.de">frank.loeffler@uni-jena.de</a>	de-RSE.org Friedrich-Schiller-Universität Jena Fürstengraben 1, 07743 Jena
Dr. Christian Grimm (angefragt) <a href="mailto:grimm@dfn.de">grimm@dfn.de</a>	Verein zur Förderung eines Deutschen Forschungsnetzes e. V. (DFN-Verein)
Prof. Dr. André Brinkmann <a href="mailto:brinkman@uni-mainz.de">brinkman@uni-mainz.de</a>	Universität Mainz
Prof. Dr. Richard Lenz <a href="mailto:richard.lenz@fau.de">richard.lenz@fau.de</a>	FAU - AG Forschungsdatenmanagement Univ. of Erlangen-Nuremberg
Prof. Dr. Ramin Yahyapour <a href="mailto:ramin.yahyapour@gwdg.de">ramin.yahyapour@gwdg.de</a>	GWDG / Uni Göttingen Am Faßberg 11, 37077 Göttingen
Prof. Albrecht Schmidt	LMU München
Prof. Klaus Pohl Direktor	paluno – The Ruhr Institute for Software Technology Universität Duisburg-Essen
Prof. Anne Koziolk	KIT Karlsruhe
Prof. Ralf Reussner	KIT Karlsruhe
Daniel Krupka, Geschäftsführer	Gesellschaft für Informatik e.V.
Prof. Dr. Thomas Ludwig	Deutsches Klimarechenzentrum DKRZ
Prof. Dr. Martin Schulz	TU München
Prof. Dr. Michael Resch	HLRS Stuttgart
Prof. Dr. Wolfgang Karl	KIT Karlsruhe
Prof. Dr. Christian Plessl	Paderborn University
Prof. Dr. Wolfgang Nagel	TU Dresden
Prof. Dr. Nico Gauger	TU Kaiserslautern
Prof. Dr. Christian Becker	Universität Mannheim
Prof. Dr. Olaf Spinczyk	Universität Osnabrück
Prof. Dr. Timo Kehler	HU Berlin
Prof. Dr. rer. nat. Vera G. Meister	TH Brandenburg