



Kurzinformation NFDI und Konsortien

Stand: März 2023

Inhaltsverzeichnis

Die Nationale Forschungsdateninfrastruktur	2
Ziel und Aufbau von NFDI	2
Historie	3
Konsortien	3
Base4NFDI	4
BERD@NFDI	5
DAPHNE4NFDI	6
DataPLANT	7
FAIRagro	8
FAIRmat	9
GHGA	10
KonsortSWD	11
MaRDI	12
NFDI4Biodiversity	13
NFDI4BIOIMAGE	14
NFDI4Cat	15
NFDI4Chem	16
NFDI4Culture	17
NFDI4DataScience	18
NFDI4Earth	19
NFDI4Energy	20
NFDI4Health	21
NFDI4Immuno	22
NFDI4Ing	23
NFDI4Memory	24
NFDI4Microbiota	25
NFDI4Objects	26
NFDI-MatWerk	27
NFDIxCS	28
PUNCH4NFDI	29
Text+	30
Sektionen	31

Die Nationale Forschungsdateninfrastruktur

Ziel und Aufbau von NFDI

Forschungsdaten sind ein wertvoller Schatz für unsere Gesellschaft. Mit ihrer Hilfe lassen sich neue Erkenntnisse gewinnen, die Innovationen ermöglichen und Deutschland als Wissensstandort stärken. Aktuell sind viele Forschungsdaten weder auffindbar noch interoperabel nutzbar und damit nicht wieder- und weiterverwertbar. Die Nationale Forschungsdateninfrastruktur (NFDI) möchte dies ändern und neue Potenziale zur Wissensgenerierung und zur interdisziplinären Verknüpfung erschließen.

Der Verein Nationale Forschungsdaten- infrastruktur (NFDI) e. V.

Gründungsmitglieder: Bund und alle 16 Länder

Gründungsdatum: 12. Oktober 2020

Mitglieder: über 225 im Januar 2023
(juristische Personen)

Vorstand: Prof. Dr. York Sure-Vetter (Direktor) und
Eva Lübke (Kaufmännische Leiterin)

Standort: Geschäftsstelle in Karlsruhe

Zweck:

- Förderung von Wissenschaft und Forschung durch eine Nationale Forschungsdateninfrastruktur
- Entwicklung eines übergreifenden FDM
- Steigerung der Effizienz der Wissenschaft

Aufgaben:

- Steuerung und Koordination beim Aufbau einer vernetzten Informationsinfrastruktur
- Schaffung eines verlässlichen Angebots von datenbasierten Diensten
- Schaffung von Daten-Standards
- Modellierung von wissenschaftsspezifischen Metadaten und Ontologien
- Klärung rechtlicher sowie ethischer Fragen

Bund und Länder haben die Chancen eines modernen Forschungsdatenmanagements (FDM) erkannt und fördern NFDI bis einschließlich 2028 jährlich mit bis zu 90 Millionen Euro.

In der Digitalstrategie Deutschland wird NFDI eine zentrale Rolle für Fortschritt in der Forschung zugeschrieben: „In diesem Datenraum wollen wir Informationen sammeln, ordnen und verfügbar machen.“

Unter dem Dach von NFDI sollen sämtliche wissenschaftlichen Fachgebiete in Konsortien, Zusammenschlüsse verschiedener Einrichtungen innerhalb eines Forschungsfeldes, vereint werden. Universitäten und Forschungseinrichtungen in ganz Deutschland arbeiten zusammen daran, Standards für ihre Communities und darüber hinaus zu etablieren. Dieser einzigartige Bottom-up-Prozess soll dafür sorgen, dass Lösungen geschaffen werden, die wirklich Anklang finden und in der Breite kompatibel sind. Relevante Daten sollen nach den FAIR-Prinzipien (Findable, Accessible, Interoperable und Reusable – also auffindbar, zugänglich, interoperabel und wiederverwendbar) zur Verfügung gestellt werden. Um die Aktivitäten

zum Aufbau einer Nationalen Forschungsdateninfrastruktur zu koordinieren, wurde der gemeinnützige Verein Nationale Forschungsdateninfrastruktur (NFDI) e.V. mit Sitz in Karlsruhe gegründet.

NFDI gestaltet ein effektives Forschungsdatenmanagement. Dies stärkt den Austausch verschiedener Disziplinen und den Wissenschaftsstandort Deutschland insgesamt. Die Vernetzung in NFDI ermöglicht einen systematischen Umgang mit Big Data sowie den Einsatz von Künstlicher Intelligenz. Zudem wird das verantwortungsvolle Teilen von Daten sowohl zwischen Wissenschaft und Wirtschaft als auch auf internationaler Ebene vereinfacht.

NFDI vertritt unter anderem als mandatiertes Mitglied Deutschland bei der Initiative European Open Science Cloud (EOSC), die einen cloudbasierten FAIRen Datenraum aufbauen möchte.

In dem ebenfalls nach Europa ausgerichteten Drittmittelprojekt FAIR Data Spaces, das vom BMBF im Zeitraum von Mai 2021 bis Mai 2024 gefördert wird, geht es darum, die föderierte, sichere Dateninfrastruktur von Gaia-X und NFDI zu einem gemeinsamen, cloud-basierten Datenraum für Industrie und Forschung unter Einhaltung der FAIR-Prinzipien zu verbinden. Ein weiteres Drittmittelprojekt, „Datenkompetenz in NFDI“, hat zum Ziel, den ganzheitlichen und nachhaltigen Aufbau von Datenkompetenz für Wissenschaftler:innen unterstützen. Es ist 2022 gestartet und wird für zwei Jahre vom BMBF finanziert.

Historie

Die Empfehlung, eine koordinierte Forschungsdateninfrastruktur in Deutschland aufzubauen, kam vom Rat für Informationsinfrastrukturen (RfII). Um die Bedürfnisse und Ansprüche verschiedener Forschungscommunitys an eine Informationsinfrastruktur möglichst gut berücksichtigen zu können, schlug der RfII die Einrichtung sogenannter Konsortien vor, in denen sich Wissenschaftler:innen nach Fachrichtungen zusammenschließen können, um passgenaue Lösungen zu etablieren. Die Vorschläge des RfII fanden in Wissenschaftskreisen breiten Anklang. Die Gemeinsame Wissenschaftskonferenz (GWK) einigte sich auf den Aufbau einer NFDI mit der Verabschiedung der Bund-Länder-Vereinbarung vom 26. November 2018. Im Mai 2019 beschloss die GWK die Ansiedlung des NFDI-Direktorats in Karlsruhe und betraute das FIZ Karlsruhe - Leibniz-Institut für Informationsinfrastruktur GmbH sowie das Karlsruher Institut für Technologie (KIT) mit den Aufgaben der Gründungsphase. Diese umfassten den Aufbau der NFDI-Geschäftsstelle und die Überführung von NFDI in eine eigene Rechtspersönlichkeit. Am 12. Oktober 2020 wurde der NFDI-Verein in Hannover von Bund und Ländern gegründet, seit 1. März 2021 ist er organisatorisch selbstständig.

Konsortien

Zu NFDI gehören 26 Konsortien und der Verbund von Konsortien Base4NFDI. Die Konsortien wurden in einem wissenschaftsgeleiteten Verfahren, das die Deutsche Forschungsgemeinschaft (DFG) steuert, ausgewählt. Sie decken vielfältige **Wissenschaftsdisziplinen** ab: von Kultur-, über Sozial-, Geistes- und Ingenieurwissenschaften bis hin zu Lebens- und Naturwissenschaften.



Base4NFDI

Base4NFDI stellt eine einzigartige Chance für das deutsche Wissenschaftssystem dar. Mit Base4NFDI unternehmen alle Konsortien von NFDI die gemeinsame Anstrengung, zuverlässige Basisdienste zu entwickeln und anzubieten. An dem Verbund von Konsortien beteiligen sich Institutionen aus allen wichtigen Bereichen der

wissenschaftlichen Infrastrukturversorgung und aus allen großen Forschungsorganisation in Deutschland. Durch die breite Zusammenarbeit von Wissenschaftsbereichen und Infrastrukturanbietern sollen Synergien in der wissenschaftlichen Dateninfrastruktur identifiziert und genutzt werden. NFDI-weite Basisdienste haben das Potential, die meisten oder alle Konsortien zu bedienen und damit einen signifikanten Einfluss auf die Effizienz der deutschen Forschungslandschaft zu haben. Unter einem "Service" wird in Base4NFDI eine technisch-organisatorische Lösung verstanden, die typischerweise Speicher- und Rechendienste, Software, Prozesse und Workflows sowie die notwendige personelle Unterstützung für verschiedene Service Desks umfasst.

Basisdienste für NFDI

Koordinator: Prof. Dr. Lars Bernard

Koordinierende Institution: Technische Universität Dresden

Gefördert seit: März 2023

Mehr Informationen: <https://base4nfdi.de/>

Ziele

- Base4NFDI bietet einen Rahmen für die nutzerorientierte und qualitätsgesicherte Entwicklung von Basisdiensten und baut dabei auf bestehende Dienste auf.
- Base4NFDI gewährleistet eine nahtlose Integration von entwickelten Basisdiensten in die NFDI-Strukturen, indem es auf die Entscheidungsprozesse des NFDI-Vereins aufbaut.
- Base4NFDI entwickelt langfristige Betriebsmodelle für NFDI-weite Basisdienste.
- Base4NFDI leistet einen wichtigen Beitrag zu internationalen Forschungsdateninfrastrukturen und -initiativen und wird eine Vorreiterrolle in der internationalen Entwicklung einnehmen.

Use Case: Identity- und Access-Management

Identity- und Access-Management (IAM) befasst sich mit den Prozessen, Richtlinien und Technologien zur Verwaltung digitaler Identitäten und deren Zugriffsrechte auf bestimmte Ressourcen. Ein zentrales Ziel innerhalb von NFDI ist es, einen einheitlichen Zugriff auf Daten, Software und Rechenressourcen sowie einen souveränen Datenaustausch und kollaboratives Arbeiten zu ermöglichen. Um dies zu erreichen, sollen bestehende und neu entstehende IAM-Systeme so verbunden und erweitert werden, dass Forschende aus verschiedenen Bereichen und Institutionen so einfach wie möglich auf digitale Ressourcen innerhalb der NFDI zugreifen können, einschließlich des Zugangs zu und des Austauschs mit externen Infrastrukturen und Ressourcen. Um die hierfür notwendige Interoperabilität zu erreichen, ist ein dezentrales, föderiertes Identitäts- und Zugangsmanagement erforderlich. Der technische und organisatorische Rahmen für ein föderiertes IAM ist eine sogenannte Authentifizierungs- und Autorisierungsinfrastruktur (AAI). In Base4NFDI wird eine dem Stand der Technik entsprechende AAI aufgebaut und bereitgestellt, die eine konsortienübergreifende und internationale Zusammenarbeit ermöglicht.

BERD@NFDI

BERD@NFDI widmet sich dem Aufbau einer leistungsstarken Plattform für die Sammlung, (Vor-)Verarbeitung, Analyse und Aufbewahrung von Wirtschafts- und verwandten Daten. Damit soll das integrierte Management von Daten und Algorithmen über den gesamten Forschungszyklus hinweg ermöglicht werden. Ein besonderer Schwerpunkt liegt auf unstrukturierten Daten, wie Video-, Bild-, Audio-, Text- oder mobilen Daten sowie Big Data.

Business, Economic and Related Data (BERD)

Sprecher: Prof. Dr. Florian Stahl

Antragstellende Institution: Universität Mannheim

Gefördert seit: Oktober 2021

Mehr Informationen: <https://www.berd-nfdi.de/>

Mit der BERD@NFDI-Plattform schafft NFDI Infrastrukturen, um den Herausforderungen des erweiterten empirischen Modells für unstrukturierte Daten zu begegnen. Das Konsortium baut eine Community auf, bietet öffentlich verfügbare und online zugängliche Datensätze an und entwickelt Standards für die Datendokumentation und -aufbewahrung gemäß der FAIR-Prinzipien. Darüber hinaus stellt BERD@NFDI ein Algorithmen-Repository, Benchmarks, Rechen- und Speicherleistung für die Analyse von (großen) Daten sowie eine breite Palette von APIs für die Interaktion mit externen Systemen bereit.

Ziele

- FAIRes Speichern, Teilen, Aufbewahren und Verknüpfen von unstrukturierten Daten aus nicht-standardisierten Quellen und von Algorithmen
- Mitwirken an der Entwicklung von Qualitäts- und Dokumentationsstandards sowie von Tools und Services für das (Vor-)Verarbeiten und Analysieren unstrukturierter Forschungsdaten für eine bessere Reproduzierbarkeit von Forschungsergebnissen
- Verbesserung der Auffindbarkeit von und des Zugriffs auf unstrukturierte Daten und Algorithmen
- Support für Forschende für das Management unstrukturierter Forschungsdaten über den gesamten Forschungsdatenzyklus hinweg

Use Case: Modernes Forschungsdatenmanagement für die Wirtschaftswissenschaften

BERD@NFDI ist eine Initiative zum Aufbau einer leistungsstarken Plattform für die Sammlung, Verarbeitung, Analyse und Aufbewahrung von Geschäfts-, Wirtschafts- und verwandten Daten. Ein Teilprojekt von BERD@NFDI befasst sich mit der semantischen Vernetzung von Unternehmensinformationen und deren Bereitstellung in einem Knowledge Graph für die Wissenschaft. Bei dem Knowledge Graph handelt es sich um eine Datenbank, in der Informationen - beispielsweise zum Standort, zur Beschäftigtenzahl oder zu den Rechtsverhältnissen eines Unternehmens, strukturiert aufgearbeitet sind, sodass daraus kompaktes Wissen entsteht. Die erste Version ist bereits auf der Konsortiumswebseite zu finden: <https://www.berd-nfdi.de/servicetools/knowledge-graph/>

DAPHNE4NFDI

DAPHNE4NFDI baut eine offene und FAIRe Infrastruktur für das Forschungsdatenmanagement für die Forschung mit Photonen und Neutronen auf. Synchrotronlicht (besonders intensive Lichtquelle mit wohldefinierbaren Eigenschaften wie z. B. Wellenlänge) und Neutronen erlauben einzigartige Einblicke in Materie sowie biologische, chemische und

physikalische Prozesse. Mit den gewonnenen Erkenntnissen können dringend benötigte Techniken entwickelt werden, die zur Lösung der aktuellen Probleme unserer Gesellschaft beitragen. Die Vision von DAPHNE4NFDI ist, die Transparenz, die FAIRness und die Wiederverwendbarkeit von Daten, die an Photonen- und Neutronenanlagen generiert werden, zu verbessern und diese Daten für die Gesamtheit der Wissenschaftler:innen zur Verfügung zu stellen. Zudem besteht großer Bedarf an schneller und fortschrittlicher Datenanalyse sowie der Nutzung von Referenzdaten, die auch zentral bereitgestellt werden. Die Daten sollen systematisch gespeichert und auf Qualität geprüft werden. Mit dieser Systematik können zusätzliche Informationen und Erkenntnisse aus den Daten extrahiert werden.

Daten aus PHotonen und Neutronen Experimenten

Sprecher: Dr. Anton Barty

Antragstellende Institution: Deutsches Elektronen-Synchrotron (DESY)

Gefördert seit: Oktober 2021

Mehr Informationen: <https://www.daphne4nfdi.de>

Ziele

- Verbesserung der Aufnahme von Metadaten, um Daten von Experimenten an Neutronen- und Photonenquellen in größerer Breite besser aufzufinden und wiederzuverwenden;
- Implementierung von Datenbanken: Repositorien für verarbeitete Daten, veröffentlichte und unveröffentlichte Ergebnisse, sowie Aufbau neuer Referenzdatenbanken;
- Entwicklung, Pflege und Bereitstellung der von den Nutzenden entwickelten Software, so dass die Datenanalyse wiederholbar und für weitere Fragestellungen wiederverwendbar ist;
- Beteiligung an einer multidisziplinären Datenplattform für Konsortien-übergreifende NFDI-Aktivitäten;
- Aus- und Weiterbildung im Bereich des Forschungsdatenmanagements für Photonen- und Neutronenexperimente.

Use Case: Dynamik in korrelierten Elektronensystemen, katalytische Materialien

Zur Erforschung neuer Materialien, z.B. für Spin-Elektronik, und für das Verständnis von Materialeigenschaften, wie z.B. der verlustfreien Leitung von Strom oder der magnetischen Kühlung, werden Neutronen und Photonen auf Materialien gesendet und ihr Energieverlust gemessen. Ganz ähnlich ist die Nutzung von Synchrotronstrahlung und Neutronen zur Erforschung von Katalysatoren, welche in ihrer Grundfunktion chemische Reaktionen beschleunigen. Neben Streumethoden werden spektroskopische Techniken genutzt. Diese Experimente sind nur an Großforschungszentren mit begrenzter Messzeit möglich; daher sollten die Daten für möglichst viele Forschende zugänglich sein. Oft werden für die Analyse verschiedene Vergleichsexperimente benötigt. In DAPHNE4NFDI werden (Referenz-)Datenbanken zu Messungen zur Struktur der Materialien vor, nach und während eines Prozesses aufgebaut und elektronische Laborbücher etabliert oder optimiert. DAPHNE4NFDI entwickelt zudem entsprechende Qualitätskontrollen und User Interfaces.

DataPLANT

Gemeinsam mit anderen Fachgebieten steht die Pflanzen-Grundlagenforschung vor der Herausforderung, ein Forschungsdatenmanagement zu etablieren, das sowohl den Bedürfnissen der Community entspricht als auch FAIR Forschungsdatensätze schafft. Diesem Ziel widmet sich DataPLANT, indem Datenqualität,

Interoperabilität und Standardisierung von Daten vorangetrieben werden. Beginnend mit der Unterstützung von weit verbreiteten Ontologien und deren Anwendung, wird eine infrastrukturelle Basis für die entwickelten Tools geschaffen. Des Weiteren dient der DataPLANT-Hub als zentraler Zugangspunkt und bestärkt die Forschenden durch Data Stewards in deren Ausbildung und Koordination des Community-Kontakts.

Pflanzen-Grundlagenforschung

Sprecher: Dr. Dirk von Suchodoletz

Antragstellende Institution: Albert-Ludwigs-Universität Freiburg

Gefördert seit: Oktober 2020

Mehr Informationen: <https://www.nfdi4plants.org/>

Ziele

- Aufbau des Annotated Research Context (ARC)
- ARC wird den gesamten Forschungszyklus abdecken, vom Experiment über die rechnerischen Aspekte bis hin zu den eigentlichen Daten und Metadaten sowie die daraus resultierenden Publikationen.
- ARC basiert auf bestehenden Formaten, Terminologien und Richtlinien, sodass die Integration bestehender Daten in bereits existierende Repositorien möglichst reibungslos erfolgen kann.

Use Case: Transdisziplinär forschen durch ARCs und DataHUB

Die Auswirkungen von Umweltveränderungen auf die Pflanzen- und Tierwelt in einem bestimmten Gebiet können Biolog:innen nur mit Expert:innen aus verschiedenen Bereichen, wie z.B. der Genetik, Ökologie und Bioinformatik gemeinsam erforschen. Für diese Kollaboration nutzen sie die durch DataPLANT angebotene Plattform. Dort kann er in Form des FAIR Digital Objects namens [ARC \(Annotated Research Context\)](#) Forschungsdaten, Metadaten und Bearbeitungsschritte in einer vorgegebenen Struktur verständlich darstellen. Alle Forschenden nutzen den [DataHUB](#), um ihre Daten zu teilen und gemeinsam zu bearbeiten; dieser basiert auf Gitlab, sodass eine standardmäßig Versionierungsfunktion die Nachverfolgung der Datenherkunft erlaubt. Das Tool [Swate](#) unterstützt dabei die im ARC strukturierten Daten mit Metadaten zu versehen, indem es sich domänenspezifischen Ontologien bedient und damit sicherstellt, dass die Daten für Menschen und Computer gleichermaßen verständlich sind. Um die akute Ontologie-Lücke zu überbrücken, können zusätzlich neue Begriffe als Term hinzugefügt und direkt genutzt werden. DataHUBs befähigen Forschende ihre Daten in Form eines ARCs auf verständliche Weise zu präsentieren und innerhalb dieser Struktur zu kollaborieren, ohne die Herkunft der Daten zu verlieren, was zu einer verbesserten Forschungsqualität und -ergebnissen führt. Schließlich plant der Biologe oder die Biologin, die Ergebnisse der gemeinsamen Forschung zu veröffentlichen und dadurch mit einem persistenten Identifikator zu versehen, was in Kürze für valide ARCs innerhalb des DataHUBs ermöglicht werden soll.

FAIRagro

FAIRagro vereint die Gemeinschaft der Agrosystemforschung und entwickelt eine maßgeschneiderte, digitale Infrastruktur. Die Agrosystemforschung schafft Grundlagen für eine nachhaltige Pflanzenproduktion in bestehenden und zukünftigen Agrarökosystemen. FAIRagro unterstützt die Agrosystemforschung, indem verschiedene Disziplinen, Skalen und Methoden durch maßgeschneidertes Forschungsdatenmanagement (FDM) verbunden werden. Weil FAIRes Datenmanagement von der Akzeptanz durch die Nutzer lebt, werden bei FAIRagro interdisziplinäre Forschende in Entwicklungsprozesse mit einbezogen. Um Forschungsdaten der Agrosystemforschung FAIR zu machen, wird eine interoperable und skalierbare Forschungsdateninfrastruktur eingerichtet. Dabei werden bestehende Repositorien miteinander vernetzt und auf diese Weise übergreifende, systemische Analysen erleichtert.

FAIRagro

Sprecher: Prof. Dr. Frank Ewert

Antragstellende Institution: Leibniz-Zentrum für Agrarlandschaftsforschung (ZALF) e.V.

Gefördert seit: März 2023

Mehr Informationen: <https://fairagro.net>

Ziele

- Einrichtung des FAIRagro-Portals als zentraler Zugangspunkt für FDM in den Agrosystemwissenschaften, um Wissensaustausch, Technologietransfer und partizipative Prozesse zur Förderung des Dialogs mit der Community zu ermöglichen.
- Einrichtung eines Data Steward Service Center, das Unterstützung im Bereich FDM anbietet, Leitlinien und Informationsmaterial entwickelt und auf einer zentralen Plattform bereitstellt.
- Kontinuierliche Identifizierung von FDM-Herausforderungen in der Agrosystem-Community und Bereitstellung von Lösungen durch agil entwickelte FDM-Dienste.
- Sicherstellung der Qualität von Forschungsdaten durch domänenspezifische Maßnahmen der Qualitätskontrolle und Einrichtung eines Qualitätsfeedback- und Kurationsystems.
- Schaffung von Rechtssicherheit für Infrastrukturanbieter, Datenanbieter und Nutzer bei der Veröffentlichung und Wiederverwendung sensibler Daten.

Use Case: Förderung des integrierten Pflanzenschutzes

Der Pflanzenschutz zielt darauf ab, Ertragseinbußen aufgrund von Schädlingen und Krankheiten zu minimieren. Es gibt jedoch zunehmend wissenschaftliche und öffentliche Bedenken hinsichtlich des Einsatzes von Pestiziden, weshalb die EU-Strategie anstrebt, den Einsatz von Pestiziden bis 2030 um 50 % zu reduzieren. Die integrierte Schädlingsbekämpfung (IPM) zielt darauf ab, den Einsatz von Pestiziden und die damit verbundenen Umweltauswirkungen zu minimieren, indem vielseitige Optionen für die Pflanzenbewirtschaftung eingesetzt werden. Trotz zunehmender Bemühungen von Politik, Wissenschaft und Beratung um die Förderung von IPM konnte bisher kein durchschlagender Erfolg erzielt werden. Ein Hauptgrund liegt in der mangelnden Auffindbarkeit, Standardisierung, Zugänglichkeit und Integration von IPM-bezogenen Daten und Modellen. Es fehlt eine integrierte Entscheidungshilfe für den Pflanzenschutz, die den potenziellen Ertragsverlust und das Umweltrisiko des Pestizideinsatzes berücksichtigt. Die künftige Integration verschiedener Modelltypen ist daher von entscheidender Bedeutung, um die Entscheidungshilfe im Bereich des Pflanzenschutzes voranzutreiben und den Pflanzenschutz zum Erfolg zu führen. .

FAIRmat

Forschungsdaten von FAIRmat sind für den Wohlstand und Lebensstil unserer Gesellschaft von maßgeblicher Bedeutung, denn Bereiche wie Energie, Umwelt, Gesundheit, Mobilität, IT usw. hängen von besseren oder neuartigen Werkstoffen ab. Ohne eine effiziente Infrastruktur sind diese Daten weder zugänglich noch wiederverwendbar. FAIRmat wird für die Physik der kondensierten Materie und die chemische Physik fester Stoffe realisieren sich der Aufgabe widmen, eine FAIRe Dateninfrastruktur (DI) zu entwickeln. Mit einer DI, die Synthese, Experiment, Theorie und Simulationen integriert, kann FAIRmat auch auf andere Bereiche wie Chemie, Ingenieurwesen, Industrie und die Gesellschaft aktiv zugehen.

Materialwissenschaften

Sprecher: Prof. Dr. Claudia Draxl

Antragstellende Institution: Humboldt-Universität zu Berlin

Gefördert seit: Oktober 2021

Mehr Informationen: <https://www.fairmat-nfdi.eu/fairmat/>

FAIRmat basiert auf umfassender Erfahrung mit der weltweit größten Dateninfrastruktur der computergestützten Materialwissenschaften, dem Novel Materials Discovery (NOMAD) Laboratory sowie dem Verein FAIR-DI e.V.. Das Konsortium ist international bestens eingebettet, unter anderem in die RDA, die EOSC, GO FAIR, etc. und hat mit führenden Institutionen wie NIST (USA), Shanghai University (China) und CSC (Finnland) Memoranda of Understanding unterzeichnet. Dementsprechend setzt sich FAIRmat für zunehmende Akzeptanz einer Forschungsdateninfrastruktur ein – in Deutschland, Europa und darüber hinaus.

Ziele

- Wissenschaftler:innen davon überzeugen, auch solche Daten zu teilen, die sich für ihre eigenen fokussierten Forschungsprojekte als nicht relevant erweisen;
- auf die erweiterte Community zugehen und umfassende Beratung, Schulungen und Unterstützung anbieten;
- eine föderierte, FAIRe DI mit zentralem Zugang – dem FAIRmat Portal – aufbauen;
- Metadatenkataloge und Ontologien (Systeme von Informationen mit logischen Relationen) entwickeln;
- den effizienten Austausch von Forschungsdaten ermöglichen und sicherstellen, dass die DI die Grundlagenforschung unterstützt, ohne dass dies zu einer Belastung für die aktiven Forscher:innen führt. Auch andere Gebieten, z.B. die Ingenieurwissenschaften, werden von der FAIRmat DI profitieren.

Use Case: Neue Materialien mit Datensätzen auffindig machen

Das Konsortium FAIRmat kombiniert erstmals Forschungsdaten aus der Synthese, experimentellen Untersuchung und Simulation von Materialien, um nutzbare Materialien zu finden, die sonst bei individueller Betrachtung verborgen blieben. Die Software, die in FAIRmat entwickelt wird, wird bereits von über einem Dutzend Laboren aus unterschiedlichen Bereichen der Materialwissenschaften (Synthese, Experiment, Simulation) in Deutschland und darüber hinaus erprobt. Aus den theoretischen Materialwissenschaften sind bereits die Ergebnisse von über zwölf Millionen Simulationen im NOMAD Laboratory öffentlich verfügbar: <https://nomad-lab.eu/nomad-lab/>

GHGA

Die Erhebung von Genomdaten von Patient:innen ist eine Schlüsseltechnologie in der biomedizinischen Forschung. Diese Art von Daten wird zunehmend auch in der Diagnostik und im Rahmen von personalisierten Therapien erzeugt. So entstehen aktuell an einer wachsenden Anzahl von Forschungsstandorten

German Human Genome-Phenome Archive

Sprecher: Prof. Dr. Oliver Stegle

Antragstellende Institution: Deutsches Krebsforschungszentrum

Gefördert seit: Oktober 2020

Mehr Informationen: <https://www.ghga.de/>

in Deutschland umfangreich charakterisierte Datensätze. Für eine optimale Nutzung dieser Daten in der Forschung ist eine weitreichende Integration dieser lokalen Datensätze unerlässlich. Nur so können moderne Computermethoden wie maschinelles Lernen oder künstliche Intelligenz ihr volles Potential entfalten. Dabei muss das Bedürfnis, Daten offen und FAIR für die Forschungsgemeinde zu erschließen immer mit dem Schutz der Privatsphäre der Patient:innen abgewogen werden. Um diesen Balanceakt zu ermöglichen, baut GHGA eine nationale Infrastruktur für humane Genomdaten auf. Diese soll es erlauben, hochsensible Genomdaten in einem einheitlichen, datenschutzkonformen Rahmen zusammenzuführen, zu speichern und zu analysieren. Als nationaler Knoten des föderierten Europäischen Genom-Phänom-Archivs (EGA) kann GHGA spezifische nationale Regelungen zum Datenschutz folgen und gleichzeitig eng an internationale Dateninfrastrukturen angebunden sein.

Ziele

- Bereitstellung eines nationalen Langzeitarchivs humaner Omics-Daten (molekularbiologische Daten);
- Adressierung rechtlicher und ethischer Hindernisse für Datenaustausch in der Forschung
- Erhöhung der FAIRness von Omics-Daten (Einbettung in nationale und internationale Dateninfrastrukturen);
- Entwicklung und Bereitstellung maßgeschneiderter Datenportale, um spezifischen Anforderungen an Datensätzen und Analysetools gerecht zu werden;
- Sensibilisierung für die effiziente und verantwortungsvollen Nutzung und Verwaltung von Omics-Daten;
- Demokratisierung des Zugriffs und der Analyse umfangreicher Omics-Daten über eine Cloud-basierte Analyseplattform.

Use Case: Mit Genomdaten Krebs besser behandeln

Für eine optimale Nutzung von Patientendaten ist eine weitreichende Integration dieser lokalen Datensätze in große Kohorten unerlässlich. Nur so können Methoden wie maschinelles Lernen oder künstliche Intelligenz sinnvoll angewandt werden.

GHGA wird eine nationale Dateninfrastruktur aufbauen, die es erlaubt Genomdaten unter Wahrung der Datenschutzauflagen zusammenzuführen und zu analysieren. In der Onkologie wird es den Ärzten zum Beispiel möglich sein, beobachtete tumorspezifische genetische Veränderungen bei ihren Patient:innen mit ähnlichen Fällen in der GHGA-Datenbank zu vergleichen.

Bei der Diagnose von seltenen genetischen Erkrankungen können Genomprofile, die im Rahmen von molekulargenetischen Untersuchungen erstellt wurden, mit den Datensätzen in GHGA verglichen werden. Kausale Zusammenhänge zwischen seltenen genetischen Erkrankungen und den ursächlichen genetischen Veränderungen können so mit Hilfe von GHGA identifiziert werden.

KonsortSWD

SWD - Strengthen, Widen, Deepen: so versteht das Konsortium für die Sozial-, Bildungs-, Verhaltens- und Wirtschaftswissenschaften seine Aufgabe innerhalb der NFDI: Die bestehende Infrastruktur für Forschungsdatenmanagement auszubauen und zu integrieren. Um gesellschaftlich relevante, insbesondere sensible Daten für die Forschung verfügbar zu

Sozial-, Bildungs-, Verhaltens- und Wirtschaftswissenschaften

Sprecher: Prof. Dr. Christof Wolf

Antragstellende Institution: GESIS – Leibniz-Institut für Sozialwissenschaften

Gefördert seit: Oktober 2020

Mehr Informationen: <https://www.konsortswd.de/>

machen, baut KonsortSWD auf bewährte wissenschaftsnahe Strukturen, insbesondere den Forschungsdatenzentren, auf. Eine zentrale Rolle spielt der Rat für Sozial- und Wirtschaftsdaten (RatSWD), der als gewähltes Gremium die Anliegen der Fachcommunitys artikuliert und so die nutzenorientierte Weiterentwicklung der Infrastruktur forciert. Die Stärke unserer Dateninfrastruktur ist das geteilte Verständnis, dass viele Daten nur dann überhaupt für die (Sekundär-)Nutzung bereitgestellt werden können, wenn sie bei den Datenanbieter:innen verbleiben. Diese sichern mit ihrer Expertise zugleich die höchstmögliche Qualität der Daten.

Ziele

- die Forschungsdateninfrastruktur zur Beforschung der Gesellschaft stärken, erweitern und vertiefen;
- Forschende und Forschungsdatenzentren technisch und inhaltlich bei Verwaltung und Nachnutzung (neuer) sensibler und nicht sensibler Daten gemäß FAIR-Prinzipien unterstützen.

Use Case: Sicheres Datenteilen auch bei sensiblen Informationen

KonsortSWD baut auf einer Dateninfrastruktur auf, die seit über 20 Jahren nachweisbar wissenschaftliche Erkenntnisse fördert. Die vom RatSWD akkreditierten, derzeit 42 Forschungsdatenzentren (FDZ) stellen amtliche sowie in der Wissenschaft erzeugte Daten zur Verfügung. Was lässt sich z.B. über gesundheitliche Faktoren und deren Auswirkung auf die Erwerbstätigkeit und das Risiko von Arbeitslosigkeit sagen? Wie wandeln sich die Einstellungen gegenüber Werten wie Gleichheit oder Toleranz und warum? Welche Faktoren erklären Bildungserfolg – in der Schule und im Arbeitsleben? Auch wenn die zugrundeliegenden Daten (Gesundheit, Beschäftigungshistorie, persönliche Einstellungen) zumeist sensitive Information enthalten, können FDZ die Zugangsmodalitäten flexibel anpassen und so adäquaten Datenschutz ermöglichen. 2021 verzeichneten die FDZ 52.758 externe Datennutzende und führten zu 3.101 wissenschaftlichen Publikationen, die mit diesen Daten arbeiteten.

MaRDI

Motiviert durch die Bedarfe und Anforderungen sowohl aus der mathematischen Forschung als auch aus wissenschaftlichen Disziplinen, die Mathematik verwenden, will das NFDI-Konsortium MaRDI Standards und Services für den Umgang mit mathematischen Forschungsdaten entwickeln und etablieren. Mathematische Forschungsdaten

reichen von Datenbanken für spezielle Funktionen und mathematische Objekte über Aspekte des Wissenschaftlichen Rechnens wie Modelle und Algorithmen bis hin zu Daten, die in statistischen Analysen anfallen oder durch Unsicherheiten gekennzeichnet sind. Durch die Interdisziplinarität und Abstraktionskraft der Mathematik sind sie auch in anderen Wissenschaftsgebieten weit verbreitet. Beim Umgang mit Forschungsdaten in der Mathematik stellt sich die Frage nach der Reproduzierbarkeit und der Wiederverwendbarkeit der Ergebnisse sowie der lokalen oder zeitlich begrenzten Verfügbarkeit der Daten. Trotz der Existenz spezieller Antworten und Lösungen fehlt eine umfassende Datenkultur und entsprechende Infrastruktur für Forschungsdaten in der Mathematik, die den Forschungsprozess unterstützen und die FAIR-Prinzipien umsetzen. MaRDI wird, ausgehend von den Bereichen Computer Algebra, Wissenschaftliches Rechnen, Statistik und maschinelles Lernen sowie der interdisziplinären Mathematik, Standards für reproduzierbare Workflows zur Berechnung zertifizierter mathematischer Ergebnisse entwickeln und neue Services für die mathematische Forschung bis zum Peer-Review zur Verfügung stellen.

Mathematische Forschungsdateninitiative

Sprecher: Prof. Dr. Michael Hintermüller

Antragstellende Institution: Weierstraß Institut für Angewandte Analysis und Stochastik

Gefördert seit: Oktober 2021

Mehr Informationen: <https://www.mardi4nfdi.de/>

Ziele

- Entwicklung der Interoperabilität für mathematische Forschungsdaten;
- Standardisierung semantischer Beschreibungen für mathematische Objekte, Algorithmen, Modelle, Software und weitere Daten zum interdisziplinären Austausch;
- Zertifizierung der Ergebnisse von exakten oder numerischen Berechnungen sowie Computerexperimenten;
- Etablierung reproduzierbarer Workflows für computerunterstützte Mathematik, Wissenschaftliches Rechnen, Statistik und maschinelles Lernen;
- Aufbau des MaRDI-Portals für mathematische Forschungsdaten über dezentrale und vernetzte Wissens- und Datenspeicher;

Use Case: Algorithmen helfen bei der Erstellung komplexer Modelle

Mathematische Algorithmen spielen eine zentrale Rolle zur Lösung praktischer Fragestellungen, z.B. in der Medizin oder in der Modellierung des Klimawandels. Die Wissenschaftler:innen, die diese Algorithmen entwickeln oder verwenden, benötigen einen Zugang zu den damit verbundenen Daten und dem Wissen. Das sind z.B. Artikel, in denen die Algorithmen diskutiert werden, Benchmarks, Software oder Informationen, welche Algorithmen ein bestimmtes mathematisches Problem lösen. Die MaRDI-Services werden über ein Portal zugänglich gemacht, das unter anderem eine gezielte Suche nach Algorithmen für bestimmte Problemklassen und den zugehörigen Forschungsdaten sowie vielen anderen ermöglicht.

NFDI4Biodiversity

NFDI4Biodiversity ist ein Netzwerk von Akteur:innen, die eines eint: das Erheben von und Arbeiten mit Biodiversitätsdaten. Zu den Partner:innen des Konsortiums zählen neben Forschungseinrichtungen auch Verbände, naturkundliche Fachgesellschaften, Landesämter, zivilgesellschaftliche Initiativen und andere

Biodiversität, Ökologie und Umweltdaten

Sprecher: Prof. Dr. Frank Oliver Glöckner

Antragstellende Institution: MARUM – Zentrum für Marine Umweltwissenschaften, Universität Bremen

Gefördert seit: Oktober 2020

Mehr Informationen: <https://www.nfdi4biodiversity.org>

Einrichtungen aus dem Bereich des Naturschutzes. Ziel ist es, eine gemeinschaftliche Vorgehensweise zur Verfügbarmachung von Biodiversitätsdaten für Forschung und Praxis zu erarbeiten. Biodiversität meint dabei mehr als nur die Erfassung der Vielfalt der Arten; sie umfasst die genotypische, phänotypische und funktionelle Diversität sowie Interaktion von Arten, Populationen und Ökosystemen. Es sind entsprechend vielfältige Daten, die erschlossen werden müssen, um die drängenden Fragen zu den Veränderungen der belebten Umwelt beantworten zu können. Hier setzt NFDI4Biodiversity an: Das Konsortium bietet seinen Nutzer:innen Dienste, die es erlauben, Daten besser zu sichern, einrichtungsübergreifend zu organisieren, zu publizieren – und somit nachnutzbar zu machen.

Ziele

- Ein leistungsfähiges Praktiker-Netzwerk schaffen, das vielfältig nutzbare Lösungen entwickelt;
- die Auffindbarkeit, Zugänglichkeit, Verknüpfbarkeit und Nachnutzbarkeit von Daten verbessern (FAIR Data); Qualität in der Datenorganisation sicherstellen;
- NFDI4Biodiversity in verwandte nationale und internationale Aktivitäten integrieren;
- die NFDI als Serviceinfrastruktur gemeinsam steuern und dynamisch gestalten;
- Forschungsdatenmanagement zum integralen Bestandteil der Biodiversitätsforschung machen.

Use Case: Monitor der Siedlungs- und Freiraumentwicklung (IÖR-Monitor)

Der IÖR-Monitor basiert auf der kombinierten Verarbeitung amtlicher Geobasis-, Geofach- und statistischer Daten. Aus diesen werden Indikatoren abgeleitet, die in thematische Kategorien wie zum Beispiel Siedlung, Freiraum oder Verkehr unterteilt sind. In einem Kartenviewer können so Landnutzung, Landbedeckung und Landschaftsqualität sowie deren Veränderung räumlich, zeitlich und thematisch hochauflösend für Deutschland dargestellt werden. Die Daten des IÖR-Monitors differenzieren 39 Bodennutzungsarten und sind für eine Reihe von Struktureinheiten – wie Bundesländern, Gemeinden oder Raumordnungsregionen – verfügbar. Unter anderem sind die Daten Grundlage für die Beschreibung von Ökosystemzuständen und die Bewertung von Ökosystemleistungen. Ziel ist es, die IÖR-Monitor-Daten in NFDI4Biodiversity zu integrieren. Andere Biodiversitätsdaten, beispielsweise zum Vorkommen von Tier- und Pflanzenarten, ließen sich somit den Daten zu Bodentypen, Versiegelung oder Besiedlung von Landflächen verknüpfen und über verschiedene Zeitspannen ins Verhältnis setzen, um neue Erkenntnisse zu gewinnen. Eine der großen Herausforderungen in NFDI4Biodiversity besteht in der Einbringung von Daten aus der einen in eine andere Infrastruktur. Dazu ist viel Vorarbeit nötig, etwa im Bereich der Datentransformation und des NFDI-Infrastrukturaufbaus. Im Fall des IÖR-Monitors ist hier der Grundstein gelegt.

NFDI4BIOIMAGE

Bioimaging bezeichnet die Gesamtheit der Verfahren, die der Visualisierung der Strukturen und Funktionsmechanismen lebender Organismen dienen. Angesichts des massiven Volumens und der Komplexität von bildgebenden Verfahren ist ein adäquates Datenmanagement für jede Art der Analyse von Bioimaging-Daten unerlässlich, allerdings bisher kaum realisiert. Für viele Disziplinen der Lebens- und Medizinwissenschaften, die auf Bioimaging angewiesen sind, ist dies ein substantielles Hindernis. NFDI4BIOIMAGE wird als methodenzentriertes Konsortium Lösungen erarbeiten, damit Bioimaging-Daten so geteilt und wiederverwendet, wie sie akquiriert werden, über disziplinäre Grenzen hinaus. Der volle Informationsgehalt dieser Daten soll ausgeschöpft, und aus der Re-Analyse neue Erkenntnisse gewonnen werden. NFDI4BIOIMAGE will eine Infrastruktur bereitstellen, die einerseits disziplinspezifischen Anforderungen gerecht wird, andererseits an andere Datentypen und Systeme verschiedener Domänen der Wissenschaft anschließt. Die ebenfalls bestehende starke internationale Vernetzung von NFDI4BIOIMAGE stellt eine Chance für deutsche Wissenschaftler:innen dar, an internationalen Initiativen mitzuwirken, die die FAIRifizierung von Bioimaging-Daten als eine Hauptherausforderung der Medizin- und Lebenswissenschaften erkannt haben.

Research Data Management for Microscopy and BioImage Analysis

Sprecher: Prof. Dr. Stefanie Weidtkamp-Peters
Antragstellende Institution: Heinrich-Heine-Universität Düsseldorf
Gefördert seit: März 2023
Mehr Informationen: <https://nfdi4bioimage.de>

Ziele

- Speicherkonzepte und offene Repositorien: Zugangs- und Übermittlungsanforderungen, Kapazität, langfristig vs. kurzfristig, verteilt vs. zentralisiert, Datenflüsse;
- Datenintegration: Kombination anderen disziplinspezifischen Datentypen, integrative Datensatzabfragen, Interoperabilität, Schnittstellen;
- Entwicklung von Bioimage-Informatik-Tools: Bioimage-Analyse, automatisierte KI-basierte Bioimage-Analyse-Pipelines, Workflow-Management und Reproduzierbarkeit, Interoperabilität von Analyse-Tools, Statistik und Modellierung;
- Speicherkonzepte und offene Repositorien: Zugangs- und Übermittlungsanforderungen, Kapazität, langfristig vs. kurzfristig, verteilt vs. zentralisiert, Datenflüsse;
- Einrichtung eines Webportals, das Informationen, Wissen und Werkzeuge für die Verwaltung von Bioimage-Daten zusammenfasst und verbreitet.

Use Case: Möglichkeiten der Hochleistungsmikroskopie ausschöpfen

Wann ist ein Gen angeschaltet? Woher kommt ein Protein? Moderne Bildgebungsverfahren mit Hochleistungsmikroskopie erzeugen komplexe, multi-dimensionale Daten. Sie gewähren Einblick in Entwicklung und Veränderung von Organen und Geweben mit hoher räumlicher und zeitlicher Auflösung. Die Verknüpfung dieser Bildinformation mit Daten aus anderen experimentellen Ansätzen wie der Genom- und Proteom-Forschung verspricht für die medizinische und biologische Forschung einen hohen Erkenntnisgewinn. Im Arbeitsbereich „Multimodale Datenintegration“ entwickelt NFDI4BIOIMAGE ein System für die interdisziplinäre Verknüpfung verschiedener Datentypen mit Bilddaten, um diese für neue Forschungsfragen nachnutzbar zu machen.

NFDI4Cat

NFDI4Cat konzentriert sich auf die Katalysatorforschung und katalyseverwandte Wissenschaften wie Chemieingenieurwesen und Verfahrenstechnik. Die Katalyse ist ein hochkomplexes und interdisziplinäres Wissenschaftsgebiet, das die effiziente Herstellung einer Vielzahl von Produkten für

NFDI für Wissenschaften rund um Katalyse

Sprecher: Dr. Andreas Förster

Antragstellende Institution: DECHEMA Gesellschaft für Chemische Technik und Biotechnologie e.V.

Gefördert seit: Oktober 2020

Mehr Informationen: <https://nfdi4cat.org/>

verschiedene Branchen und in unterschiedlichen Produktionsmaßstäben ermöglicht. Für die Katalyse als interdisziplinäre Wissenschaft mit hohem Anwendungsbezug ist NFDI4Cat die Möglichkeit, alle Teildisziplinen in Bezug auf Datenmanagement und Datenaustausch zu integrieren und den höchstmöglichen Mehrwert zu generieren. Die Notwendigkeit einer gemeinsamen Vorgehensweise wird in der GeCatS-Publikation „The Digitalization of Catalysis-Related Sciences¹“ beschrieben. NFDI4Cat umfasst Partner aus allen Bereichen der Katalyse, so auch die Reaktions- und Verfahrenstechnik. Die Katalyse- und Ingenieurkompetenzen werden durch Expertise in den Datenwissenschaften, des Hochleistungsrechnens und des maschinellen Lernens ergänzt. Das Konsortium bündelt eine große Bandbreite an vorhandenem Know-how, Repositorien und Software-Werkzeugen, auf die aufgebaut werden kann, um eine gezielte, effiziente und fächerübergreifende Arbeit zu gewährleisten.

Ziele

- Förderung der disziplinübergreifenden Forschung im Bereich der Katalyse
- Vision „vom Molekül zum Prozess“ auf digitaler Ebene realisieren
- Brücke zwischen den verschiedenen Fachcommunitys sowie zwischen Industrie und Wissenschaft
- Klärung zentraler Fragen zu geistigem Eigentum, Zugriffsrechten und Anreizsystemen

Use Case: Moderne Verfahren für die Entwicklung von Katalysatoren

Die Entwicklung von Katalysatoren ist ein komplexer, jahrelanger Prozess mit vielen Beteiligten, sodass der Datenaustausch oft limitiert und nicht standardisiert ist. Dies reduziert die Effizienz der Forschung und erschwert, moderne Verfahren wie maschinelles Lernen bei der Entwicklung von Katalysatoren einzusetzen. Dafür werden besser auffindbare, zugängliche, interoperable und vor allem wiederverwendbare Daten benötigt. NFDI4Cat arbeitet an der Vereinheitlichung strukturierter Daten und der Standardisierung von Daten- und Metadatenformaten, inkl. Werkzeuge und Leitfäden. Durch eine gemeinsame virtuelle Infrastruktur für diese Daten wird die Zusammenarbeit zwischen akademischer und angewandter Katalyse durch informationsgetriebenes Design weiter vorangetrieben. Die Verfügbarkeit hochwertiger aktueller und vorhandener Forschungsdaten wird zudem ermöglichen, neuartige katalytische Reaktionswege in einem frühen Stadium im Hinblick auf Energieeffizienz und Nachhaltigkeit zu bewerten. Die Katalysatoren und Verfahren können synergetisch entwickelt und in den Bereichen Energiewende, Ernährung, Emissionsminderung, Energieeffizienz und Nachhaltigkeit zukunftsorientiert eingesetzt werden.

¹ [GeCatS_Whitepaper+2019_engl_ezl-p-20005018.pdf\(dechema.de\)](#)

NFDI4Chem

NFDI4Chem baut eine offene und FAIRe Infrastruktur für das Forschungsdatenmanagement in der Chemie auf. Das Konsortium besteht aus engagierten Produzent:innen und Nutzer:innen von Daten aus der universitären und außeruniversitären Forschung, aus Infrastruktureinrichtungen und aus Fachgesellschaften wie der Gesellschaft Deutscher Chemiker (GDCh), der Bunsen-Gesellschaft und der Deutschen Pharmazeutischen Gesellschaft (DPhG). NFDI4Chem möchte alle Disziplinen der Chemie im akademischen Bereich repräsentieren.

Fachkonsortium Chemie in der NFDI

Sprecher: Prof. Dr. Christoph Steinbeck

Antragstellende Institution: TIB – Leibniz-
Informationszentrum Technik und
Naturwissenschaften Universitätsbibliothek, Hannover

Gefördert seit: Oktober 2020

Mehr Informationen: <https://www.nfdi4chem.de>

Die Vision von NFDI4Chem ist es, alle wichtigen Schritte der chemischen Forschung zu digitalisieren, um Wissenschaftler:innen bei der Erfassung, Speicherung, Verarbeitung, Analyse, Offenlegung und Nachnutzung von Forschungsdaten zu unterstützen. Maßnahmen zur Förderung von Open Science und des Forschungsdatenmanagements (FDM) gemäß den FAIR-Prinzipien sind grundlegende Ziele von NFDI4Chem, um der Chemie-Community ein ganzheitliches Konzept für den Zugang zu Forschungsdaten zu bieten. Zu diesem Zweck entwickelt NFDI4Chem eine nationale Forschungsdateninfrastruktur für die Forschungsdomäne Chemie in Deutschland. Sie wird innovative und einfach zu nutzende Dienste bieten und neue wissenschaftliche Ansätze auf Basis der Nachnutzung von Forschungsdaten ermöglichen. In der Anfangsphase konzentriert sich NFDI4Chem auf Daten zu Molekülen und Reaktionen einschließlich Daten zu deren experimenteller und theoretischer Charakterisierung.

Ziele

- Virtuelle Umgebung von föderierten Repositorien für Forschungsdaten in der Chemie;
- Internationale Community-Prozesse zur Etablierung von Standards für Mindestangaben und offene Daten in der Chemie;
- Kultureller und digitaler Wandel hin zu „Smart Laboratory Environments“;
- Bewusstsein und Akzeptanz für FAIRes Datenmanagement in der chemischen Gemeinschaft;
- Förderung von Synergien und bereichsübergreifenden Entwicklungen innerhalb der NFDI;
- Rechtlich verlässlicher Rahmen von Richt- und Leitlinien für FAIRes FDM.

Use Case: Ressourcensparenden chemische Reaktionsweg leichter finden

Es wird eine vernetzte Infrastruktur aus Software-Tools, elektronischen Laborjournalen und Daten-Repositorien entwickelt und bereitgestellt. Diese wird Forschende im Laboralltag unterstützen. Möchte beispielsweise eine akademische Gruppe oder ein Industrieunternehmen einen besonders ressourcensparenden chemischen Reaktionsweg (z.B. durch neuartige Katalysatoren) für ein neues Medikament oder auch eine neue Art von Solarzelle entwickeln, dann wird es nach Etablierung der NFDI4Chem Zugriff auf Tausende von Daten zu erfolgreich getesteten Reaktionen haben. Damit lässt sich nicht nur das avisierte Entwicklungsproblem lösen. Ebenso können chemieinformatisch arbeitende Gruppen Methoden des tiefen maschinellen Lernens verwenden, um künstliche Intelligenzen zur Reaktionsplanung zu entwickeln.

NFDI4Culture

Digitale Daten zu materiellen und immateriellen Kulturgütern sind ein wesentlicher Bestandteil des täglichen Lebens, der Kommunikation und der kulturellen Erfahrung. Sie sind nicht nur für künftige Wissenschaftsgenerationen von großer Bedeutung, sondern bilden die Grundlage von beständigen Transferprozessen zwischen Forschenden, Kunstschaffenden, Kulturwirtschaft und Zivilgesellschaft. Die langfristige Bewahrung und zuverlässige Verfügbarkeit von digitalen Kulturgütern im Einklang mit den FAIR-Prinzipien sind von grundlegender Bedeutung für das kulturelle Selbstverständnis von Individuen, Gruppen und der Gesellschaft als Ganzes. Ziel von NFDI4Culture ist der Aufbau einer bedarfsorientierten Infrastruktur für Forschungsdaten zu materiellen und immateriellen Kulturgütern innerhalb der Gesamtstruktur von NFDI. NFDI4Culture zielt darauf ab, das breite Spektrum der verschiedenen Akteur:innen im Bereich des Kulturerbes in idealer Weise zu repräsentieren.

Konsortium für Forschungsdaten zu materiellen und immateriellen Kulturgütern

Sprecher: Prof. Torsten Schrade

Antragstellende Institution: Akademie der Wissenschaften und der Literatur, Mainz

Gefördert seit: Oktober 2020

Mehr Informationen: <https://nfdi4culture.de/>

Die langfristige Bewahrung und zuverlässige Verfügbarkeit von digitalen Kulturgütern im Einklang mit den FAIR-Prinzipien sind von grundlegender Bedeutung für das kulturelle Selbstverständnis von Individuen, Gruppen und der Gesellschaft als Ganzes. Ziel von NFDI4Culture ist der Aufbau einer bedarfsorientierten Infrastruktur für Forschungsdaten zu materiellen und immateriellen Kulturgütern innerhalb der Gesamtstruktur von NFDI. NFDI4Culture zielt darauf ab, das breite Spektrum der verschiedenen Akteur:innen im Bereich des Kulturerbes in idealer Weise zu repräsentieren.

Ziele

- Aufbau einer bedarfsorientierten Infrastruktur für Forschungsdaten zu materiellen und immateriellen Kulturgütern in der NFDI;
- Fächerübergreifende Auffindbarkeit und Zugänglichkeit sowie die langfristige Sicherung und kontinuierliche Pflege der Forschungsdaten;
- Dazu gehören 2D-Digitalisate von Gemälden, Photographien und Zeichnungen ebenso wie digitale 3D-Modelle kulturhistorisch bedeutender Gebäude, Denkmäler oder audiovisuelle Daten von Musik-, Film- und Bühnenaufführungen;
- Kontinuierliche Zusammenarbeit auf Augenhöhe, die bedarfsgerechte Ausrichtung aller Dienste und der regelmäßige Transfer innovativer Forschungsmethoden und -ergebnisse in die Fachcommunities, die NFDI, die Kulturpolitik, die Kulturwirtschaft und die interessierte Zivilgesellschaft.

Use Case: Saving Ukrainian Cultural Heritage Online

Wie essenziell der dauerhafte Erhalt und der digitale Zugang zu materiellen und immateriellen Kulturgütern für die Bewahrung und Fortschreibung kultureller Identitäten ist, zeigt sich aktuell im Krieg in der Ukraine. Mitwirkende von NFDI4Culture haben sich bereits wenige Tage nach Kriegsausbruch in der internationalen Initiative „Saving Ukrainian Cultural Heritage Online“ (SUCHO) an der digitalen Rettung ukrainischer Kulturgüter beteiligt. NFDI4Culture stellt im Rahmen einer Task Force mehrere Dienste zur Verfügung: Neben einem spezifischen Ukraine-Helpdesk zur Meldung und Vermittlung von Digitalisierungsbedarfen ukrainischer Kultureinrichtungen gehört dazu auch die Bereitstellung von Expertise zur Klärung technischer, ethischer und rechtlicher Herausforderungen. Durch die Kooperation mit SUCHO wirkt NFDI4Culture dem drohenden Wissens- und Kulturverlust in der Ukraine entgegen. Die in diesem Zusammenhang entwickelten Workflows und das Know-how des Konsortiums stehen auch für denkbare künftige Szenarien im In- und Ausland bereit.

NFDI4DataScience

Die Vision von NFDI4DataScience (NFDI4DS) ist es, alle Schritte des komplexen Lebenszyklus von Forschungsdaten in Datenwissenschaften und Künstlicher Intelligenz zu unterstützen. In den letzten Jahren hat sich ein Paradigmenwechsel vollzogen: Berechnungsmethoden stützen sich zunehmend auf Daten- und oft auch Deep-Learning-basierte Ansätze, was dazu geführt hat, dass sich die Datenwissenschaften als eine Disziplin etabliert hat, die von Fortschritten im Bereich der Informatik angetrieben wird. Transparenz, Reproduzierbarkeit und Fairness sind aufgrund der Komplexität der datenwissenschaftlichen Methoden, die oft auf einer Kombination aus Code, Modellen und Daten beruhen, zu entscheidenden Herausforderungen für Datenwissenschaften und Künstliche Intelligenz geworden.

Das übergreifende Ziel von NFDI4DS ist die Entwicklung, der Aufbau und die Erhaltung einer nationalen Forschungsdateninfrastruktur für die Forschungsgemeinschaft der Datenwissenschaften und Künstlichen Intelligenz. In der Anfangsphase wird sich NFDI4DS auf vier Anwendungsbereiche konzentrieren: Sprachtechnologie, biomedizinische Forschung, Informationswissenschaften und Sozialwissenschaften.

Ziele

- Erhöhung von Transparenz, Reproduzierbarkeit und Fairness von Projekten;
- Unterstützung aller Schritte des komplexen und interdisziplinären Lebenszyklus von Forschungsdaten;
- Verfügbar machen aller digitalen Artefakte (wie Code, Modelle, Daten oder Publikationen), Verknüpfung dieser und Angebot zusätzlicher Werkzeuge und Dienste;
- Repräsentation der Forschungsgemeinschaft in Datenwissenschaften und Künstlicher Intelligenz in Deutschland.

Use Case: Daten, Modelle, Publikationen und Skripte verknüpfen

Gesichtserkennungs-Algorithmen haben zum aktuellen Zeitpunkt typischerweise Probleme bei der Unterscheidung von Menschen mit dunkler Hautfarbe. Dies ist in der Regel auf die verwendeten Trainingsdaten zurückzuführen. Um dieser und ähnlichen Herausforderungen zu begegnen, stellt NFDI4DS eine offene Forschungsdateninfrastruktur bereit, die es ermöglicht, neben Publikationen und Daten auch Modelle und Skripte verfügbar zu machen, und diese miteinander zu verknüpfen. Die Infrastruktur setzt auf diverse bereits etablierte Dienste auf, wie etwa <https://dblp.org>.

NFDI für Datenwissenschaften und Künstliche Intelligenz

Sprecher: Dr. Sonja Schimmler

Antragstellende Institution: Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.

Gefördert seit: Oktober 2021

Mehr Informationen:

<https://www.nfdi4datascience.de/>

NFDI4Earth

NFDI4Earth stellt ein umfassendes Netzwerk mit Hochschulen, Großforschungseinrichtungen, Ressortforschung sowie Fachgesellschaften und Vereinen dar und die Wissenschaftler:innen kooperieren in internationalen und interdisziplinären Netzwerken. Ihr Ziel ist es, die Funktionsweisen und Wechselwirkungen im

Erdsystem zu verstehen und die vielfältigen Herausforderungen des globalen Wandels anzugehen. Diese sind z.B. die vielfältigen disziplinspezifischen Ansätze, die Ableitung gemeinsamer Standards und die Unterstützung der Interoperabilität heterogen verteilter Dienste. NFDI4Earth ist ein Community-gesteuerter Prozess, der Forschenden einen FAIRen, kohärenten und offenen Zugang zu allen relevanten Erdsystem-Daten, innovativen Forschungsdatenmanagement- und Data-Science-Methoden ermöglicht. Darüber hinaus soll ein breit akzeptiertes NFDI4Earth FAIRness and Openness Bekenntnis das Kernstück zur Förderung eines kulturellen Wandels zu FAIRem und offenem Forschungsdatenmanagement in den Erdsystemwissenschaften bilden.

Erdsystemwissenschaften

Sprecher: Prof. Dr. Lars Bernard

Antragstellende Institution: Technische Universität Dresden (TUDD)

Gefördert seit: Oktober 2021

Mehr Informationen: <https://www.nfdi4earth.de/>

Ziele

- Einfacher, effizienter, offener und, wann immer möglich, unbegrenzter Zugang zu allen relevanten Daten, Datenmanagement- und Analysediensten der Erdsystemwissenschaften;
- Bereitstellung eines zentralen Zugangs zu strukturierten Informationen und Hilfestellungen;
- Etablierung einer verteilten, nachhaltigen User-Support-Netzwerkstruktur;
- Vielfältige Unterstützung der Aus- und Weiterbildung durch Bereitstellung einsatzbereiter Bildungsressourcen und Etablierung eines Netzwerks für Nachwuchswissenschaftler:innen;
- Berücksichtigung vielversprechender neuer Tools und Trends für innovative Data-Science-Ansätze im Incubator Lab und durch Pilot Projekte;
- Aktive und kontinuierliche Einbeziehung verschiedener Stakeholder der Erdsystemwissenschaftlichen Community und disziplinübergreifende Interaktion im NFDI-Netzwerk.

Use Case: Verlässliche Klimadaten für relevante Entscheidungen

Den Klimawandel "sichtbar" zu machen, ist von hoher gesellschaftlicher Relevanz. Daher ist es wichtig, qualitätsgesicherte Klimadaten aus allen Sphären des Erdsystems transparent zugänglich zu gestalten, um belastbares Entscheidungswissen für Politik und Gesellschaft zu generieren. NFDI4Earth – als nationaler Verbund der Erdsystemwissenschaften in Deutschland – ermöglicht eine neue Qualität von FAIRen Daten, mit denen die Diagnose und Projektion des Klimawandels weiter verbessert werden kann. Darüber hinaus vereinfacht NFDI4Earth die Zugänglichkeit aller Daten über Organisationen hinweg und schafft interoperable Angebote an Forschende anderer Disziplinen, die Klimadaten benötigen (z.B. für weiterführende Planungen im Gesundheitssektor oder für die Klimaadaptation von Städten). Dies soll auch in den nächsten Jahren weiter vorangetrieben werden. Damit wird ein einzigartiger nationaler Datenraum geschaffen, der international sowohl konkurrenz- als auch anschlussfähig ist, und Klima- und Umweltdaten als Grundlage für fundierte – auch politische – Entscheidungen noch zugänglicher macht.

NFDI4Energy

Die Energiewende stellt die Energiesystemforschung vor Herausforderungen. Digitalisierungsprozesse erleichtern den Wandel hin zu cyber-physischen Energiesystemen (CPES - Verbund informatischer, softwaretechnischer Komponenten mit mechanischen und elektronischen Teilen) und wirken sich auf technische, soziale und gesellschaftliche Aspekte, aber auch auf den Forschungsprozess selbst aus. Forschung zu CPES stützt sich in hohem Maße auf modell- und (co-)simulationsbasierte Ansätze. Hierbei stellt die Nachverfolgung von Daten und Modellen eine komplexe Herausforderung dar, die in jedem Forschungsprojekt neu bearbeitet werden muss. Zusätzlich zeichnet sich die Energiesystemforschung dadurch aus, dass sowohl Industriepartner als auch die Gesellschaft mit unterschiedlichen relevanten Gruppen eingebunden werden muss. Die zusätzlichen Stakeholder steigern die Komplexität des Umgangs mit Daten und Modellen noch weiter. Um Forschende zu unterstützen, greift NFDI4Energy bestehende FDM-Ansätze, die unter anderem in anderen NFDI-Konsortien entwickelt werden, auf und passt diese an die Bedürfnisse der eigenen Forschungscommunity an.

NFDI für die interdisziplinäre Energieforschung

Sprecher: Prof. Dr. Astrid Nieße

Antragstellende Institution: Carl von Ossietzky
Universität Oldenburg

Gefördert seit: März 2023

Mehr Informationen: <https://nfdi4energy.uol.de/>

Ziele

- Entwicklung allgemeiner Forschungsdienste für die Energiesystemforschung;
- Ermöglichen der Nachvollziehbarkeit, Reproduzierbarkeit und Transparenz für Forschungssoftware und -daten in der Energiesystemforschung;
- Förderung der Einbeziehung der Gesellschaft in die Energiesystemforschung;
- Förderung einer besseren Zusammenarbeit und eines besseren Wissenstransfers zwischen wissenschaftlichen Forschungseinrichtungen und Industriepartnern;
- Vereinfachung der Identifizierung, Integration und Koordination von simulationsbasierten Modellen in der Forschung;

Use Case: Digitale Interaktion zur Verbesserung des Energiesystems

Aufgrund der zunehmenden Integration erneuerbarer Energien ist die digitale Interaktion und Koordination zwischen allen Akteuren im Energiesystem erforderlich, um die hohe Volatilität (Schwankung) der Erzeugung und des Verbrauchs auf allen Netzebenen effizient zu bewältigen. Flexibilitäten können z. B. durch Marktmechanismen angereizt werden, was eine engere Zusammenarbeit zwischen Netzbetreibern und privaten Akteuren, die Flexibilitätsoptionen auf niedrigeren Spannungsebenen anbieten können, erfordert. Selbstorganisation durch autonome Systeme, die durch Software-Agenten repräsentiert werden, sind die Grundlage für adaptive und selbstheilende Systeme, insbesondere in zukünftigen Anwendungsbereichen wie Redispatch (Eingriff zur Anpassung der Leistungseinspeisung) auf der Grundlage von Marktprinzipien. Die Forschung in diesem Bereich stützt sich auf Simulationen, verschiedene Datenquellen und eine gute Interaktion mit der Gesellschaft. Mit NFDI4Energy können Forscher:innen moderne Simulationsdienste nutzen, um bestehende Modelle zu koppeln und vorhandene Daten für ihre Forschung zu verwenden.

NFDI4Health

NFDI4Health hat zum Ziel, ein umfassendes Inventar deutscher epidemiologischer, Public-Health- und klinischer Studiendaten aufzubauen. Die Erfassung und Analyse dieser Daten sind wesentlich zur Entwicklung neuer Therapien, übergreifender Versorgungsansätze und präventiver Maßnahmen. Personenbezogene Gesundheitsdaten verlangen einen besonderen Schutz. Erklärtes Ziel von NFDI4Health ist, Sicherheit und Nutzbarkeit zu vereinen. NFDI4Health möchte neue Möglichkeiten zur wissenschaftlichen Nutzung von personenbezogenen Gesundheitsdaten unter Wahrung des Datenschutzes zu eröffnen. Basierend auf den FAIR-Prinzipien werden den Forschenden Dienste zur Verfügung gestellt, um Daten in interoperabler Form dauerhaft zu speichern, semantisch anzureichern und zu teilen sowie um Daten aus verschiedenen Quellen zusammenzuführen. NFDI4Health wird die Leistungsfähigkeit und Qualität der gesundheitswissenschaftlichen Forschung steigern und somit helfen, die Gesundheit der Bevölkerung zu verbessern.

Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten

Sprecher: Prof. Dr. Juliane Fluck

Antragstellende Institution: ZB MED – Informationszentrum Lebenswissenschaften

Gefördert seit: Oktober 2020

Mehr Informationen: <https://www.nfdi4health.de>

Ziele

- Die Auffindbarkeit von und den Zugang zu strukturierten Gesundheitsdaten aus klinischen und epidemiologischen Studien, Krankheitsregistern, administrativen Gesundheitsdatenbanken und der öffentlichen Gesundheitsüberwachung in Deutschland zu ermöglichen;
- Implementierung eines Rahmens für die zentrale Suche und den Zugang zu bestehenden dezentralen Infrastrukturen für epidemiologische und klinische Studiendaten;
- Erleichterung der gemeinsamen Nutzung von Daten, der Verknüpfung von Datensätzen, der harmonisierten Bewertung der Datenqualität und der geteilten Analyse von personenbezogenen Gesundheitsdaten;
- Ermöglichung der Entwicklung und des Einsatzes neuer, maschinell verarbeitbarer Zustimmungsmechanismen und innovativer Datenzugangsdienste;
- Förderung der Interoperabilität der derzeit fragmentierten IT-Lösungen in Bezug auf Metadaten-Repositoryn, Kohortendurchsuchung, Datenqualität und Harmonisierung.

Use Case: preVIEW – COVID-19

COVID-19 hat gezeigt, wie wichtig die rasche Verfügbarkeit fundierter Informationen ist, also dass wissenschaftliche Daten leichter auffindbar, standardisiert und verknüpfbar sind. So können Ressourcen in der Forschung gebündelt und schneller Erkenntnisse über das Virus gewonnen werden – für eine effektivere Bekämpfung der Pandemie. Die zunehmende Bedeutung von Preprints für die COVID-19-Forschung gab den Anstoß zur Entwicklung der Preprint-Suchmaschine preVIEW, eine leichtgewichtige semantische Suchmaschine, die sich auf die einfache Einbeziehung spezialisierter COVID-19 Textsammlungen konzentriert und eine benutzerfreundliche Webschnittstelle für die semantische Informationsbeschaffung bietet. Derzeit sind Krankheiten, menschliche Gene und SARS-CoV-2-Proteine annotiert, und weitere werden in Zukunft hinzugefügt. Das Tool erleichtert es der Fachcommunity, relevante Studien frühzeitig zu finden und Daten besser austauschen zu können.

NFDI4Immuno

NFDI4Immuno ist eine offene Infrastruktur, die das Ziel verfolgt, ein Forschungsdatenmanagement (FDM) gemäß der FAIR-Prinzipien mit Schwerpunkt Immunologie aufzubauen. Dabei wird langfristig angestrebt, alle für die Immunologie relevanten Datentypen in der NFDI abzubilden. Immunologische Forschung erstreckt sich über alle lebenswissenschaftlichen Betrachtungsebenen,

also von molekularen Mechanismen bis zu der Populationsebene des Wirtsorganismus. Ein sinnvolles FDM muss daher die einzelnen Ebenen so verknüpfen, dass domänen-spezifische Entitäten und Prozesse abgebildet werden können. Um eine Interoperabilität mit anderen Domänen zu gewährleisten, müssen aber auch domänenübergreifende Standards und Ontologien identifiziert und verwendet werden. Unsere Vision ist ein Netzwerk föderierter Repositorien für alle Daten, die den Zustand des Immunsystems beschreiben, zusammen mit Werkzeugen und Diensten, die integrierte Datenanalysen über diese Repositorien hinweg erleichtern. NFDI4Immuno möchte es ermöglichen, Daten und Metadaten von verschiedenen Experimentalplattformen zu kombinieren, und so eine komplementäre Betrachtungsweise erlauben.

National Research Data Infrastructure for Immunology

Sprecher: Dr. Christian Busse

Antragstellende Institution: Deutsches Krebsforschungszentrum

Gefördert seit: März 2023

Mehr Informationen: <https://www.nfdi4immuno.de>

Ziele

- Datenrepräsentationen, Metadatenstandards, Ontologien und Anwendungsschnittstellen mit anderen NFDI-Konsortien zusammen angleichen, um Suchanfragen und Querreferenzen über mehrere Konsortien hinweg zu ermöglichen;
- Einheitliche Metadatenbeschreibungen für bestehende und neue Datensätze erzeugen und sie durch Repositorien bereitstellen;
- Ein Netzwerk aus föderierten Repositorien für immunologische Daten aufbauen und betreiben;
- Werkzeuge und Dienste entwickeln, die eine standardisierte und reproduzierbare Datenanalyse ermöglichen.

Use Case: Forschung zu Immunreaktionen und therapeutischen Abwehrstoffen

Eine der geplanten Aktivitäten von NFDI4Immuno liegt im Aufbau von föderierten Datenrepositorien für Immunrezeptor-Repertoiredaten. Diese Daten ermöglichen nicht nur Rückschlüsse darüber, wie stark und wie spezifisch Untergruppen von Immunzellen durch z.B. ein Virus aktiviert wurden, als genetische Blaupause von Antikörpern und T-Zellrezeptoren stellen sie auch die Grundlage für therapeutische Abwehrstoffe und adaptive T-Zelltherapien dar. NFDI4Immuno betreibt daher öffentliche Repositorien für diese Daten, beteiligt sich aktiv an der Weiterentwicklung des hierfür von der Community verwendeten Software-Stacks und wird eigene und bereits öffentliche Daten Dritter kuratieren und veröffentlichen.

NFDI4Ing

NFDI4Ing repräsentiert die deutsche Ingenieurforschungslandschaft. Das Konsortium vereint Institutionen und Personen aus allen Bereichen ingenieurwissenschaftlicher Tätigkeiten, die ihre jeweils spezifische Expertise, Erfahrung und Netzwerke in das Arbeitsprogramm des Konsortiums einbringen. Damit ermöglicht NFDI4Ing die enge

Vernetzung der Perspektiven von Ingenieurwissenschaftler:innen, Anwender:innen und Diensteanbieter:innen, um auf das gemeinsame Ziel hinzuarbeiten, die FAIR-Prinzipien wissenschaftlichen Datenmanagements für die Ingenieurwissenschaften Realität werden zu lassen. Eine zentrale Aufgabe von NFDI4Ing ist es, die große Vielfalt ingenieurwissenschaftlicher Forschungsansätze und -methoden in einer begrenzten Anzahl gemeinsamer Standards und Anforderungen an Forschungsdatenmanagement zu konsolidieren. Drei Aspekte stehen dabei im Fokus: Bildung in Datenkompetenz möglichst von Beginn der Ausbildung bis in die Wissenschaftskarriere, Entwicklung und Erprobung eines Governance-Konzepts für den Umgang mit Forschungsdaten sowie Sicherstellung einer breiten Verfügbarkeit von Technologien und Services für maschinenlesbare Daten und deren Metadaten.

Nationale Forschungsdateninfrastruktur für die Ingenieurwissenschaften

Sprecher: Prof. Dr. Robert Schmitt

Antragstellende Institution: RWTH Aachen

Gefördert seit: Oktober 2020

Mehr Informationen: <https://nfdi4ing.de/>

Ziele

- Reproduzierbarkeit aller Schritte ingenieurtechnischer Forschungsprozesse ermöglichen;
- Forschungssoftware selbst als Forschungsdaten etablieren;
- Nachverfolgen von Proben und Datenbearbeitungsschritten automatisieren;
- Gemeinsame Nutzung und Integration von Daten über vernetzte Infrastruktur ermöglichen;
- Maschinell verarbeitbare Repräsentationen von Metadaten erzeugen;
- Verbesserung der datenbezogenen Ausbildung und Förderung des kulturellen Wandels;
- Datenpublikationen standardisieren und anerkennen.

Use Case: Digitale Zwillinge im Straßenverkehr

Die Erhebung von Felddaten, z.B. im Straßenverkehr, ist ressourcenintensiv, daher ist es besonders wichtig, die erhobenen Daten möglichst vielfältig (nach)nutzbar zu machen: Die Systeme besitzen eine Vielzahl von Sensoren und Rechnern. Mit dem Konzept des Digitalen Zwillings, einer realitätsgetreuen digitalen Abbildung eines technischen Systems bzw. einer Maschine, adressiert das Konsortium die Verknüpfung von Daten, Algorithmen und dem realen technischen System. Ein Versuchsfahrzeug für die Validierung automatisierter Fahrfunktionen zeichnet Daten mit Kameras und Laserscannern im Straßenverkehr auf und analysiert diese, um z.B. ein Ausweichmanöver einzuleiten. Zudem sendet das Fahrzeug die Daten an den Digitalen Zwilling, der die Daten speichert, analysiert und anschließend seine Modelle zur Nachbildung des realen Fahrzeugs adaptiert. Die Daten des Versuchsfahrzeugs sind Positions- und Geschwindigkeitsdaten, Kamerabild und Laserscanner-Punktwolken. Möchte man die Stelle in den Daten finden, in denen ein anderes Fahrzeug die Vorfahrt genommen hat, muss der Datensatz aufwändig händisch durchsucht werden. Damit dieser zeitintensive Prozess der Vergangenheit angehört, entwickelt NFDI4Ing auch Metadatenschemata für automatische Klassifikationsmethoden, z. B. mittels Machine Learning. Das Konzept des Digitalen Zwillings ist nicht auf technische Systeme beschränkt, sondern wird auch auf sozio-technische Systeme wie Städte angewandt.

NFDI4Memory

NFDI4Memory vereint Forschungseinrichtungen, Gedächtnisinstitutionen und Informationsinfrastrukturen aus den historisch arbeitenden Geisteswissenschaften. Ein zentrales Anliegen ist, die spezifische Tradition der Quellenkritik für das digitale Zeitalter weiterzuentwickeln.

Konsortium für historisch arbeitende Geisteswissenschaften

Sprecher: Prof. Dr. Johannes Paulmann
Antragstellende Institution: Leibniz-Institut für Europäische Geschichte
Gefördert seit: März 2023
Mehr Informationen: <https://4memory.de/>

Ohne Empfehlungen und Richtlinien fehlt Forschenden die Orientierung, wie geeignete Standards ausgewählt und eingehalten werden können. Bestehende Datenstandards sind oft fachspezifisch oder nicht breit anwendbar, was die Interoperabilität der Daten einschränkt. Benötigt wird eine bedarfsorientierte, integrierende, servicebasierte Infrastruktur, die einen kohärenten Datenraum für die Community ermöglicht. Durch fehlende Ausbildung mangelt es vielen Forschenden an den notwendigen Grundfertigkeiten, um mit digitalen Ansätzen die Arbeitsabläufe zu verbessern, neue Perspektiven aus der Analyse digitaler Daten zu gewinnen und ihre Quellenkritik mit digitalen Methoden zu erweitern. Anreize für Datenpublikationen fehlen, wodurch Forschenden in den Geisteswissenschaften signalisiert wird, dass sie ihrer Arbeit und der Karriere nicht nützen. Eine geringe institutionelle Unterstützung bedingt schließlich, dass Forschende von den notwendigen Forschungsdatenmanagement-Maßnahmen abgekoppelt bleiben. NFDI4Memory setzt hier an.

Ziele

- Sicherstellung der Qualität historischer Daten durch die Festlegung klarer Leitlinien für den Umgang mit Daten und Metadaten;
- Verbesserung der Interoperabilität und Wiederverwendbarkeit historischer Daten und Datensammlungen durch die Harmonisierung und Umsetzung klar definierter Standards;
- Ausbau der Verbindung bestehender Datensammlungen durch Erweiterung der plattformübergreifenden Suchmöglichkeiten und durch die zunehmende Bereitstellung verschiedener historischer Daten für Forschungszwecke;
- Verbesserung der digitalen Kompetenz von historisch orientierten Geisteswissenschaftler:innen im Hinblick auf die Verwaltung von Forschungsdaten;
- Förderung eines kulturellen Wandels in historisch orientierten Fachgebieten durch einen aktiven Dialog mit den entsprechenden Gemeinschaften.

Use Case: Verbindung historischer Daten

NFDI4Memory fördert die Qualität der Forschungsdaten in den historisch orientierten Geisteswissenschaften, indem es die Vielfalt der Datentypen und -formate, verschiedene Aspekte der Datenqualität, technische Überlegungen, rechtliche Bestimmungen und ethische Grundsätze berücksichtigt. Darüber hinaus erfordern historische Daten die Kontextualisierung, die das Herzstück der Geschichtswissenschaft ist. Ein digitaler Wissensraum in NFDI4Memory kontextualisiert digitale Objekte, um alle relevanten Daten miteinander zu verbinden – von analogen Quellen bis zur digitalen Präsentation der Forschungsergebnisse – und um Transparenz und Nachvollziehbarkeit zu fördern.

NFDI4Microbiota

Mikroben (Bakterien, Archaeen, einzellige Eukaryonten und Viren) spielen eine Schlüsselrolle für die Gesundheit von Mensch, Tier und ihrer Umwelt. Damit leistet die molekularbiologische Erforschung von mikrobiellen Spezies und Gemeinschaften einen wichtigen Beitrag für die Medizin, die Ökologie, die Landwirtschaft und

den Klimaschutz. Neue technologische Fortschritte in der Hochdurchsatz-Analyse führen dabei zur Generierung umfangreicher Datenmengen, deren Verwaltung, gemeinsame Nutzung, Wiederverwendung und Interpretation eines der Ziele von NFDI4Microbiota darstellt. Bestehend aus zehn erfahrenen Partnern und unterstützt durch fünf Fachgesellschaften und mehr als 50 teilnehmenden Institutionen hilft NFDI4Microbiota der Mikrobiologieforschungsgemeinschaft, indem es technische Lösungen bereitstellt, die Interaktion fördert, Schulungen anbietet und als zentrale Datendrehscheibe fungiert. Durch die Schaffung eines Cloud-basierten Systems bietet das Konsortium seinen Nutzern die Speicherung, Integration und Analyse von mikrobiellen Daten, insbesondere von Omics-Daten, an, um diese konsistent und reproduzierbar allen Bereichen der Biowissenschaften zugänglich zu machen. Damit generiert NFDI4Microbiota eine Infrastruktur, die es erlaubt, die Ergebnisse unabhängiger Projekte miteinander zu verbinden, und sorgt durch die enge Verknüpfung von Ausbildung und Services für eine effiziente und damit exzellente mikrobiologische Forschung.

NFDI für Mikrobiota-Forschung

Sprecher: Prof. Dr. Konrad Förstner

Antragstellende Institution: ZB MED –
Informationszentrum Lebenswissenschaften

Gefördert seit: Oktober 2021

Mehr Informationen: <https://www.ghga.de/>

Ziele

- Bewusstseins-schaffung für die Wichtigkeit der FAIRifizierung von Daten, Open Science und der reproduzierbaren Forschung in der mikrobiologischen Gemeinschaft;
- Reproduzierbarkeit von Forschung durch Standardisierung der Datenprozessierung und Analyse sicherstellen;
- Bereitstellung von adäquaten Methoden für die Übersetzung bestehender Daten in neue Erkenntnisse;
- Unterstützung der Gemeinschaft durch umfassende Schulungsprogramme für eine effiziente und datenbasierte mikrobiologische Forschung;
- Vernetzung der nationalen sowie internationalen mikrobiologischen Gemeinschaften.

Use Case: Forschung zu Mikroben und Lebensmittelproduktion

Die Resistenz gegen antimikrobielle Mittel ist ein großes Problem in der Lebensmittelproduktion, insbesondere im Hinblick auf die Mikrobe *Listeria monocytogenes*, deren verschiedene Stämme Berichten zufolge gegen eine Vielzahl von antimikrobiellen Substanzen resistent sind. Mithilfe der Multi-omics-Integration kann die funktionelle Reaktion einer mikrobiellen Zelle auf eine antimikrobielle Substanz untersucht werden, um zu verstehen, wie die antimikrobielle Resistenz die Zelle in die Lage versetzt, mit solchen Verbindungen in der Umwelt zu überleben. In NFDI4Microbiota werden verschiedene Daten kombiniert, um die zelluläre Reaktion von *Listeria monocytogenes*-Stämmen auf subletale Konzentrationen eines Biozids zu untersuchen, das üblicherweise in Fleischfabriken verwendet wird.

NFDI4Objects

Materielle Hinterlassenschaften sind die einzige Wissensquelle für den weit größten Teil der kulturellen und biologischen Entwicklung der Menschheit. Von Menschen gefertigte Werkzeuge, gebaute Monumente und geformte Landschaften geben Aufschluss über technologische Entwicklung, Innovation, Wechselbeziehungen zwischen Mensch und Umwelt, Migration, kulturellen Austausch und Wandel. Ihre Aussagekraft erhalten diese Überreste durch vielschichtige Analysen und kontextbezogene Informationen. Ihre Vielfalt, mannigfaltigen Kontexte und komplexen (Objekt-) Biographien stellen dabei eine besondere Herausforderung für digitale Forschungsprozesse dar. Zur Entwicklung von Lösungen hat sich mit NFDI4Objects ein Konsortium gebildet, das sich dem materiellen Erbe von rund drei Millionen Jahren Menschheits- und Umweltgeschichte widmet. Das Konsortium repräsentiert eine breite Community, die zahlreiche Fachrichtungen aus den Geistes- und Naturwissenschaften und viele unterschiedliche Institutionen wie Universitäten, außeruniversitäre Forschungseinrichtungen, Denkmalämter, Museen und Verbände, die durch gemeinsame Anforderungen an das Forschungsdatenmanagement (FDM) miteinander verbunden sind.

Forschungsdateninfrastruktur für die materiellen Hinterlassenschaften der Menschheitsgeschichte

Sprecher: Dr. Philipp von Rummel

Antragstellende Institution: Deutsches Archäologisches Institut (DAI)

Gefördert seit: März 2023

Mehr Informationen: <https://www.nfdi4objects.net/>

Ziele

- NFDI4Objects ist Impulsgeber und dient der Verbesserung von digitalen Forschungs- und Arbeitsprozessen innerhalb der Fachcommunitys. Das Konsortium tritt für die Etablierung der FAIR-Prinzipien ein und begleitet aktiv die digitale Transformation der Arbeitsmethoden.
- NFDI4Objects stellt sich der Aufgabe, große und komplexe Datenbestände aus Forschungsprozessen zu erschließen und gleichzeitig den nachhaltigen und langfristigen internationalen Zugang zu digitalen Ergebnissen von Forschungsprojekten entsprechend den Bedürfnissen der Nutzenden zu ermöglichen.
- NFDI4Objects steht für ein gemeinsames Portfolio von fachgerechter Beratung sowie zuverlässigen, bedarfsorientierten und sicheren Diensten im konsortialen Netzwerk.

Use Case: Mobile Erfassung archäologischer Funde

Die Einlieferung eines Großteils der archäologischen Kleinfunde erfolgt durch genehmigte Sondengänger:innen, Ehrenamtliche und zufällige Finder:innen. Diese Neufunde stellen eine wichtige wissenschaftliche Quelle für die Forschung dar. Die Funde werden zumeist bei Denkmalpflegeämtern physisch abgegeben und Fundmeldungen oft handschriftlich ausgefüllt. Bei der nachträglichen Überführung der Daten in bestehende Dateninfrastrukturen der Landesämter entsteht ein erheblicher Aufwand für alle Beteiligten. Im Rahmen von NFDI4Objects soll eine Infrastruktur zur mobilen Erfassung von archäologischen Kleinfunden inkl. Datenupload in die zentralen Datenbanken entwickelt werden. So sollen die Funddaten über normierte Schnittstellen für die Verarbeitung der Funde in den Funddatenbanken der Landesdenkmalämter bereitgestellt werden. Somit können die Finder nicht nur über die Fundobjekte informieren, die Daten werden auch sogleich der wissenschaftlichen Community zur Erforschung zur Verfügung gestellt.

NFDI-MatWerk

Nicht erst seit der Industrialisierung hängt unser gesellschaftlicher und wirtschaftlicher Fortschritt von der Beherrschung von Werkstoffen ab. Von Anbeginn waren und sind Materialwissenschaft und Werkstofftechnik – MatWerk – Schlüsseltechnologien ihrer Zeit. Aufgrund ihrer Vielseitigkeit halten die beiden interdisziplinären Fachgebiete eine Fülle

innovativer Lösungen bereit, um gesellschaftlichen Herausforderungen in den Zukunftsfeldern Energie, Mobilität, Umwelt etc. zu begegnen und damit einen wesentlichen Beitrag zum nachhaltigen Umgang mit unseren endlichen Ressourcen zu leisten. Eine Herausforderung stellen dabei die vielen strukturellen Skalen und damit die verschiedenen experimentellen und numerischen Methoden dar, die es in der MatWerk-Forschung abzudecken gilt. Eine vollständige Abbildung über alle Skalen hinweg ist unumgänglich, da die mechanischen und funktionellen Eigenschaften von Materialien durch ihre Mikrostruktur und damit auch durch deren Veränderungen bestimmt werden. Zu diesem Zweck strebt NFDI-MatWerk eine nahtlose Integration dezentraler Daten und Metadaten, experimenteller und numerischer Workflows und einer Materialontologie an, um die Interoperabilität und Reproduzierbarkeit der Verarbeitung von Forschungsdaten zu maximieren.

Nationale Forschungsdateninfrastruktur für Materialwissenschaften & Werkstofftechnik

Sprecher: Prof. Dr. Christoph Eberl

Antragstellende Institution: Fraunhofer Gesellschaft zur Förderung der angewandten Forschung e.V.

Gefördert seit: Oktober 2021

Mehr Informationen: <https://nfdi-matwerk.de/>

Ziele

- Entwicklung und Integration einer einheitlichen Materialontologie für eine höhere Interoperabilität zwischen heterogenen Materialdaten, Analysetools und Materialmodellen;
- Austausch und Zugriff auf Materialrohdaten und Metadaten basierend auf den FAIR-Prinzipien;
- Entwicklung eines Leitfadens für eine nutzerorientierte Materialinfrastruktur anhand von Participant Projects (PP) und der sich daraus ableitenden Infrastruktur-Nutzungsfälle (IUCs);
- Gemeinsame Nutzung von Werkzeugen und modularisierten Workflows für die experimentelle, theoretische und datengesteuerte Materialwissenschaft;
- Einbindung der Community in den Aufbau der Forschungsdateninfrastruktur sowie deren nachhaltige Nutzung.

Use Case: Analyse von Datennutzungsprofilen

Die Entwicklung einer Datenbank-Infrastruktur versteht das MatWerk-Konsortium als einen Community-getriebenen Prozess. Dazu werden Datennutzungsprofile von Participant Projects (PP) aus verschiedenen Unterdisziplinen analysiert, um die relevantesten wissenschaftlichen Szenarien innerhalb von MatWerk zu identifizieren. Die daraus resultierenden Infrastructure Use Cases (IUCs) helfen bei der kontinuierlichen Entwicklung und Überprüfung der Infrastruktur und damit auch der Validierung zu erreichender Teilziele im Einklang mit den Bedarfen unserer Fach-Community. Die verschiedenen IUCs liefern forschungsrelevante Nutzungsprofile mit realen Daten und Workflows, die realitätsnahe Tests während und am Ende von Entwicklungsphasen ermöglichen. Auf dieser Basis implementiert NFDI-MatWerk (a) eine Qualitätssicherung für die Implementierung und den Einsatz, (b) eine ständige Ausrichtung aller Aktivitäten an übergeordneten Zielen und (c) eine klare Zukunftsperspektive für die nächsten Entwicklungsschritte der Forschungsdateninfrastruktur.

NFDIXCS

Das Kernziel von NFDIXCS ist die Schaffung einer Infrastruktur, die es ermöglicht, sogenannte Forschungsdatenmanagement-Container (RDMC) zu erstellen, zu speichern, zu finden und weiterzuentwickeln. Diese Container dienen als Zeitkapsel, um die Forschungsdaten eines Projekts oder einer Publikation und die verschiedenen anderen Informationen, wie z.B.

Software, Ausführungsumgebung und zusätzliche Kontextinformationen, miteinander zu verknüpfen. Dies ermöglicht die effektive Rekonstruktion der Forschungsdaten auch Jahre später zur Reproduktion und/oder Wiederverwendung. Die entwickelten Dienste werden für einen nachhaltigen Betrieb ausgelegt. Dies geht einher mit einer gezielten Community-Entwicklung, um einen kulturellen Wandel voranzutreiben, den akademischen Nachwuchs zu fördern, Praktiken und Normen auszuhandeln und die Kompetenzen für einen verantwortungsvollen Umgang mit Forschungsdaten zu entwickeln.

Nationale Forschungsdateninfrastruktur für und mit Computer Science

Sprecher: Prof. Dr. Michael Goedicke

Antragstellende Institution: Universität Duisburg-Essen

Gefördert seit: März 2023

Mehr Informationen: <https://nfdixcs.org/>

Ziele

- Etablierung von Standards, Verfahren und Leitlinien für verschiedenen Datentypen;
- Schaffung zuverlässiger und interoperabler Datenverwaltungsdienste;
- Förderung der Anwendung von Informatikmethoden wie Big Data, Künstliche Intelligenz und Maschinelles Lernen;
- Verbesserung der Reproduktion von Forschungsergebnissen und fächerübergreifende Wiederverwendung von Daten.

Use Case: Transformationen und Filter für Daten

Oft werden Forschungsdaten durch Messungen im Zeitverlauf erhoben. Typischerweise müssen Personen eine bestimmte Aufgabe ausführen, um gegebene Hypothesen bezüglich verschiedener Interaktionstechniken zu testen. Die Daten werden vom System (z. B. Zeit, in der eine Oberfläche berührt wird) und durch Messung von Parametern der Benutzer:innen (z. B. Herzfrequenz) erfasst. Die Art dieser Daten in solchen Studien wirft eine Reihe von Aspekten bei der Veröffentlichung der Daten auf. Um nur drei zu nennen: 1) Die Daten stehen in engem Zusammenhang mit der Experimentierumgebung. Dazu gehören Software, integrierte Messinstrumente oder die verwendeten Analysewerkzeuge und die Ausführungsumgebungen wie Betriebssystem und Bibliotheken. Der RDMC-Ansatz kapselt die (Meta-) Daten zusammen mit diesen Komponenten, um sie als eine Einheit zu veröffentlichen und so eine Reproduzierbarkeit zu ermöglichen. 2) Die Einbeziehung von Menschen und die Erfassung ihrer Bewegungs- oder physiologischen Daten erfordert einen gewissen Aufwand zur Wahrung der Privatsphäre (z. B. Anonymisierung der Daten). Der erforderliche Aufwand hängt von der konkreten Studie, dem experimentellen Design sowie dem Nutzungsszenario für diese Daten ab. Der RDMC-Ansatz ist mit spezifischen Plugins und Workflow-Modellen ausgestattet, um Transformationen und Filter auf den Daten zu implementieren. Drittens kann die entwickelte Technologie mit bestimmten geistigen Eigentumsrechten verbunden sein. Die Autor:innen können sich dafür entscheiden, nur Teile zu veröffentlichen, die mit einer bestimmten Lizenz verbunden sind, die die entsprechenden Filter und Transformationen des Datensatzes für die Veröffentlichung und Wiederverwendung festlegt.

PUNCH4NFDI

PUNCH4NFDI vereint die Expertise im Datenmanagement von Teilchenphysik, Astroteilchenphysik, Hadronen- und Kernphysik sowie Astronomie. Das Konsortium kann vor allem im Bereich von skalierbaren Lösungen für sehr große Datenmengen in heterogenen Umgebungen, für viele Tausend User und mit „offenen Daten“ auf einen in Jahrzehnten

gewachsenen Erfahrungsschatz seiner Communities zurückgreifen. Die bereits vorhandenen, erfolgreichen, aber oft isoliert voneinander existierenden Lösungsansätze will PUNCH4NFDI bündeln und weiterentwickeln. Basierend auf einem in Teilen der Community bereits implementierten Schichtenmodell des Datenmanagements will das Konsortium eine FAIRe “Science Data Platform” als Biotop für den gesamten Lebenszyklus von digitalen Forschungsprodukten schaffen und auch anderen Wissenschaftszweigen zur Verfügung stellen. Ein weiteres Thema ist das Problem der Daten-Irreversibilität, also der aufgrund der sehr hohen Datenmengen unvermeidlichen Reduktion von Datensätzen. Besondere Priorität hat für PUNCH4NFDI die Vernetzung und das Etablieren von Synergien mit anderen Konsortien aus dem Bereich der Physik und darüber hinaus. Weitere Schwerpunkte sind Ausbildung und Training.

Teilchen, Universum, Kerne und Hadronen für NFDI

Sprecher: Dr. Thomas Schoerner-Sadenius
Antragstellende Institution: Deutsches Elektronen-Synchrotron DESY
Gefördert seit: Oktober 2021
Mehr Informationen: <https://www.punch4nfdi.de/>

Ziele

- Erstellung einer anschlussfähigen Definition von digitalen Forschungsprodukten;
- Aufbau des Prototyps einer “Science Data Platform” SDP als Biotop für Forschungsprodukte;
- Bereitstellung einer möglichst breiten Palette an Werkzeugen zur Datenmanipulation und -transformation in der SDP;
- Bereitstellung der notwendigen Speicher- und Computing-Technologien, basierend auf Cloud-Lösungen und der Nutzung heterogener Ressourcen (Stichwort “data lakes”);
- Lösungen für das Irreversibilitätsproblem auf Zeitskalen von Nanosekunden bis Minuten.

Use Case: Suche nach dunkler Materie

Die dunkle Materie stellt aktuell eines der größten Rätsel der Physik dar; nach ihr wird in unterschiedlichen Experimenten gesucht: an Beschleunigern, in kryogenen Untergrundexperimenten oder auch mit Satelliten (bspw. LHC, Xenon1T, PLANCK). Die Daten dieser Experimente haben unterschiedliche Formate und ihr Verständnis erfordert viele zunächst inkompatible Software-Pakete. Die in der “PUNCH Science data platform” (SDP) lebenden digitalen „Research Products“ beinhalten die Daten und die für ihre Analyse notwendigen Software-Pakete inklusive aller Schnittstellen und ermöglichen so die vereinfachte Generierung neuen Wissens. Die Abbildung der Workflows in der SDP erlaubt darüber die einfache Variation und Überprüfung von Datenanalysen oder Daten-Theorie-Vergleichen, beispielsweise auch im Rahmen des Peer-Review-Prozesses.

Text+

Der Verbund Text+ wird text- und sprachbasierte Forschungsdaten langfristig erhalten und ihre breite Nutzung in der Wissenschaft ermöglichen. Die Text+ Infrastruktur konzentriert sich zunächst auf die Datendomänen digitale Sammlungen, lexikalische Ressourcen und Editionen. Diese sind von hoher Relevanz für alle sprach- und textbasierten Disziplinen, speziell für Sprachwissenschaften, Literaturwissenschaften, Philosophie, Klassische Philologie, Anthropologie, außereuropäische Kulturen und Sprachen sowie sprach- und textbasierte Forschung der Sozial-, Wirtschafts-, Politik-, und Geschichtswissenschaften. Sie sind mit methodologischen Paradigmen verknüpft, die jeweils charakteristische, aber auch bereichsübergreifende Praktiken der Datenerzeugung, -nutzung, -analyse, -vernetzung und -kuratierung erfordern. Die Nutzungsfreundlichkeit der Daten und Werkzeuge, verbunden mit der rechtskonformen Nutzung von Sprach- und Textdaten, die häufig Rechte Dritter berührt, sind zentrale Themen von Text+. Hinzu kommt die Diversität der Daten, die bezüglich Sprachräumen (auch über Europa hinaus) und Modalitäten von Sprache und Schriftsystemen heterogen sind.

Sprach- und textbasierte Forschungsdateninfrastruktur

Sprecher: Prof. Dr. Erhard Hinrichs

Antragstellende Institution: Leibniz Institut für Deutsche Sprache Mannheim

Gefördert seit: Oktober 2021

Mehr Informationen: <https://www.text-plus.org/>

Ziele

- Entwicklung einer ortsverteilten Infrastruktur für Sprach- und Textdaten;
- Unterstützung von Forschenden bei der Erstellung, Nachnutzung, Erhaltung von Sprach- und Textdaten;
- Ermöglichung innovativer Forschung durch leichten Zugang zu Forschungsdaten und Werkzeugen.

Use Case: Von Sprach- und Textdaten zu komplexen Systemen

Sprach- und Textdaten sind in der Forschung und Industrie allgegenwärtig. Plakative Beispiele für vermarktete Anwendungen sind aus dem Bereich der Sprachtechnologie mit Spracherkennung etwa Siri oder Alexa, im Bereich der Sprachsynthese "Der ICE 1024 hat 20 Minuten Verspätung, Grund ist eine Störung im Betriebsablauf". Sprachmodelle, wie sie von ChatGPT verwendet werden, automatische Übersetzung wie Deep-L oder Google-Translate basieren auf qualitative hochwertigen Sprachaufnahmen und großen linguistisch aufbereitete Textsammlungen, die etwa im Bereich des maschinellen Lernens angewandt werden. In den Philologien werden mittels Methoden wie dem „Distant Reading“ digital vorliegende Daten in großen Mengen untersucht, z. B. um Ähnlichkeiten zwischen Texten zu untersuchen – einschließlich von Fragen, ob ein Werk eher dem Stil von Goethe oder Schiller zuzuordnen ist, der Stilometrie. Sprach- und Textdaten werden für die Grundlagenforschung zu diesen Technologien und zu neuen wissenschaftlichen Fragestellungen benötigt. Die Forschung kooperiert daher nicht nur mit Verlagen, Bibliotheken, Autor:innen, sondern auch mit Technologieunternehmen, von sprachtechnologischen Start-ups bis zu großen Unternehmen, wie Suchmaschinenbetreibern. Für den Austausch und die Nutzung der Daten können vielfältige Normen genutzt und entwickelt werden, dies geschieht zusammen mit Organisationen wie DIN, ISO oder W3C.

Sektionen

Die Konsortien arbeiten auch über die Grenzen ihrer Disziplinen hinweg zusammen. Viele Themen sind für alle oder mehrere Konsortien relevant. Diese Querschnittsthemen werden gemeinsam identifiziert und in Sektionen, also thematischen Abteilungen, bearbeitet.

Seit Oktober 2021 arbeiten Vereinsmitglieder in vier Sektionen zusammen. Die Sektion Industry Engagement ist außerdem im Entstehungsprozess.

Common Infrastructures

Ziel ist die Schaffung einer multi-cloud-basierten Infrastruktur, die von den Konsortien zum Austausch und zur gemeinsamen Nutzung von Daten-, Software- und Computer-Ressourcen genutzt werden kann.

Ethical, Legal and Social Aspects

Die Sektion bearbeitet ethische, rechtliche und soziale Fragestellungen rund um das Datenteilen. Themen sind u.a. Datenschutz oder Rechte des geistigen Eigentums. Entwickelt werden Lösungsansätze, Leitlinien und Stellungnahmen zu aktuellen (Gesetzes-)entwürfen.

(Meta-)daten, Terminologien, Provenienz

Damit Forschungsdaten gefunden und effizient nachgenutzt werden können, müssen diese mit Metadaten beschrieben werden. Metadaten sind Informationen über die Daten, z.B. über die Lizenz, unter der die Daten genutzt werden dürfen. Die Sektion unterstützt Prozesse der Konsortien bei der Erarbeitung und Etablierung geteilter Daten- und Metadatenstandards.

Training and Education

Datenkompetenz ermöglicht, die Qualität der Ressource "Daten" zu steigern und Erkenntnisgewinne zu beschleunigen. Sie soll z.B. Studierenden, Forschenden oder Personen in der Wirtschaft vermittelt werden. Dazu entwickelt und bündelt die Sektion Angebote zur Aus- und Weiterbildung.

Industry Engagement

Die Sektion befindet sich im Gründungsprozess und wurde noch nicht final im Verein eingerichtet. Sie wird sich mit der Gestaltung von Kooperationen zwischen Wirtschaft und NFDI befassen. Ziel ist, durch eine koordinierte Zusammenarbeit neue Mehrwerte zu erschließen.

Nationale Forschungsdaten- infrastruktur (NFDI) e.V.

Albert-Nestler-Straße 13
76131 Karlsruhe

+49 721 988 994 0

info@nfdi.de
www.nfdi.de

