

FAIRmat

FAIRe Dateninfrastruktur für die Materialwissenschaften und verwandte Forschungsgebiete

Konsortium für die Bewerbung bei der Nationalen Forschungsdateninfrastruktur (NFDI)

Koordination:

Claudia Draxl (HU Berlin), Matthias Scheffler und Carsten Baldauf (FHI Berlin)

Forschungsdaten sind ein wichtiger Rohstoff des 21. Jahrhunderts. Das Thema hat viele Facetten und beinhaltet die Prozessierung, Beschreibung (Metadaten), Aufbereitung, Kuratierung, mittel- und langfristige Speicherung, Kompression, Software und Hardware, sowie die Nachnutzung der Daten etc. Große Datenmengen (Big Data) ermöglichen neue und neuartige Forschungsmethoden – von der direkten Datennutzung bis hin zur Datenanalyse mit unterschiedlichsten Methoden der künstlichen Intelligenz. In den Materialwissenschaften wird die datengetriebene Forschung auch als viertes Paradigma bezeichnet.¹ In diesem Sinne werden die Offenlegung von Forschungsdaten der vielfältigen Forschergruppen und Institute und die Möglichkeit eines *effizienten* Datenaustauschs die wissenschaftliche Forschung massiv voranbringen.

Auf Empfehlung des Rates für Informationsinfrastrukturen (RfII) sollen in den kommenden Jahren in Deutschland die dringend benötigten, nachhaltigen Infrastrukturen für die Bereitstellung, Vernetzung und Pflege von Forschungsdaten geschaffen werden. Um diese Nationale Forschungsdateninfrastruktur (NFDI) in die Praxis umzusetzen, können sich Konsortien aus den unterschiedlichsten Wissenschaftssparten bewerben. Die DFG wird die fachliche Begutachtung und Bewertung übernehmen;² die Gemeinsame Wissenschaftskonferenz (GWK) von Bund und Ländern wird die Förderentscheidung treffen.³ Die Finanzierung dieses langfristigen Programms erfolgt über das BMBF, wobei für die nächsten 10 Jahre Fördermittel von bis zu 90 Mio Euro pro Jahr zur Verfügung stehen sollen. Informationen zum Ablauf des Prozesses wurden von der DFG kürzlich veröffentlicht (<http://www.dfg.de/foerderung/programme/nfdi/>).

FAIRmat (<https://fairdi.eu/fairmat>) vertritt insbesondere die Interessen der experimentellen, theoretischen und computergestützten Festkörperphysik und Materialwissenschaften. Dies schließt z.B. auch die chemische Physik fester Stoffe und die Katalyse mit ein. Ziel von FAIRmat ist es, die für die Weiterentwicklung dieser Wissenschaftsdisziplinen existierenden Infrastrukturen zu verbessern, auszubauen und miteinander zu vernetzen und – wo nötig – neue Infrastrukturen zu schaffen. Dies wird in enger Kooperation mit Computer- und Datenzentren und Bibliotheken geschehen. Die

¹ C. Draxl and M. Scheffler, Big-Data-Driven Materials Science and its FAIR Data Infrastructure. Plenary Chapter in Handbook of Materials Modeling (eds. S. Yip and W. Andreoni), Springer (2019). Download preprint: <https://fairdi.eu/asdf>

² DFG Pressemitteilung vom 7.12.2018 (<https://bit.ly/2sohAm5>)

³ <https://www.dfg.de/foerderung/programme/nfdi/index.html>

Zusammenarbeit zwischen universitären und außeruniversitären Forschungseinrichtungen sowie der Industrie wird genutzt und intensiviert. FAIRmat wird all dies innerhalb der NFDI *user-driven* und unbürokratisch umsetzen, damit die entwickelten Werkzeuge der Wissenschaft helfen und von den im jeweiligen Feld aktiven WissenschaftlerInnen akzeptiert werden.

FAIRmat bildet die *materialwissenschaftlichen* Säulen des Vereins FAIR-DI e.V. (FAIR Data Infrastructure for Physics, Chemistry, Materials Science, and Astronomy e.V., <https://fairdi.eu>) ab. In diesem Sinne sind die Ziele von FAIRmat, Dateninfrastrukturen nach den FAIR Principles^{1,4} zu schaffen, d.h. Daten sollen auffindbar (*findable*), zugänglich (*accessible*), interoperabel (*interoperable*) und nachnutzbar (*re-usable*) – kurz FAIR – sein. FAIRmat hat somit bereits umfangreiche praktische Erfahrung und verfolgt auch FAIR-DI e.V.'s Mission Statement:

Scientific data are a significant raw material of the 21st century. To exploit its value, a proper infrastructure that makes it Findable, Accessible, Interoperable, and Re-purposable – FAIR – is a must. For the fields of computational and experimental materials science, chemistry, and astronomy, FAIRdi sets out to make this happen. This enabling of extensive data sharing and collaborations in data-driven sciences (including artificial intelligence tools) will advance basic science and engineering, reaching out to industry and society.

FAIRdi bringt in FAIRmat *umfangreiche, praktische* Erfahrungen aus dem Bereich der Computational Materials Science ein. Basierend auf dem Projekt NOMAD (*Novel Materials Discovery*, <https://nomad-coe.eu/>) und den damit verbundenen Aktivitäten betreibt NOMAD innerhalb von FAIR-DI e.V. das weltweit größte Datenrepositorium dieser Wissenschaftsdisziplin und hat auch die zugehörigen Metadaten und Infrastrukturen entwickelt. Die Komponente NOMAD ist auch als *Implementation Network* von GO FAIR⁵ aktiv. Aufbauend auf diesen Erfahrungen wird in FAIRmat das bereits Erreichte in Hinblick auf die Interessen der experimentellen und chemischen Physik der kondensierten Materie, funktioneller Materialien und der Katalyse weiterentwickelt – eine enorme Herausforderung und Chance zugleich.

FAIRmat legt großen Wert darauf, dass die zu entwickelnde Dateninfrastruktur und die dazugehörigen Metadatenstruktur **von den aktiven WissenschaftlerInnen und dem breiten Benutzerkreis** konzipiert und entwickelt werden. Die **Akzeptanz** des Feldes und eine **unbürokratische** Teilnahme aller interessierter Personen und Institutionen sind dabei sehr wichtig. Deshalb organisiert FAIRmat auch entsprechende Workshops, z.B. zur Weiterentwicklung und Verknüpfung von Metadaten.⁶ Anders als bei dem erwähnten „Erfolgsmodell“ NOMAD, wird es bei den nächsten Schritten nicht grundsätzlich darum gehen, zentrale Repositorien für das gesamte Feld zu erstellen. Es müssen verteilte Lösungen gefunden und umgesetzt werden, die den Ansprüchen des jeweiligen Experiments und/oder Forschungsbereichs genügen.

⁴ M. D. Wilkinson et al. *The FAIR Guiding Principles for scientific data management and stewardship*, Sci Data 3, 160018 (2016).

⁵ GO FAIR: a bottom-up international approach for the practical implementation of the European Open Science Cloud (EOSC) as part of a global Internet of FAIR Data & Services, <https://www.go-fair.org/>

⁶ Metadata workshop, organisiert von FAIRdi.eu, Juli 2019 <https://th.fhi-berlin.mpg.de/meetings/meta2019>

Um diese Aufgaben zu meistern und damit die Wissenschaft voranzubringen, wollen wir möglichst viele KollegInnen in FAIRmat integrieren. Viele von Ihnen kämpfen bereits heute mit Big Data als Problem, und fast alle haben erkannt, dass dies in wenigen Jahren auch ihre Arbeit stark beeinflussen wird. Gemeinsam wollen wir mit FAIRmat dazu beitragen, Grundlagenforschung, angewandte Forschung und Industrie in die Lage zu versetzen, die dringenden Herausforderungen unseres Forschungsgebietes und unserer Gesellschaft effizienter als bisher zu lösen (Energiewende, Ressourcenknappheit, Umweltverschmutzung, Medizintechnik, Kommunikation, Mobilität sind nur einige Themen).

Auch wenn sich FAIRmat als deutsches Konsortium versteht, ist der internationale Austausch mit Initiativen wie der European Open Science Cloud (EOSC) von großer Bedeutung, um Insellösungen zu vermeiden. Natürlich plant FAIRmat eine enge Abstimmung und Zusammenarbeit mit anderen NFDI-Initiativen, wie zum Beispiel NFDI4Chem und NFDI4Ing.

Die konkrete Struktur von FAIRmat befindet sich noch in der Entwicklung. In einem ersten Ansatz unterscheiden wir nach Methoden und Forschungsrichtungen, weil in diesen sehr unterschiedliche Anforderungen zu Tage treten. Die Themen beinhalten auf theoretischer Seite etwa computerunterstützte Materialwissenschaft, Simulationen weicher Materie (Biomoleküle, Polymere), auf experimenteller Seite Elektronenmikroskopie, Spektroskopie (inkl. Synchrotrons) oder die Forschungsfelder Oberflächenphysik, Metallphysik, Halbleiterphysik, Chemische Physik Fester Stoffe, Katalyse, Weiche Materialien, etc. Weitere Vorschläge werden gerne aufgenommen.



















Zurzeit gibt es fünf Aufgabenbereiche, in denen sich verschiedene Untergruppen bilden werden:

- **Computational Materials Science** [Claudia Draxl (HU Berlin), Matthias Scheffler (FHI Berlin), Silvana Botti (FSU Jena), et al.]: Electronic-structure theory, (*ab initio*) molecular dynamics, Green-function based methods, wave-function based methods, quantum chemistry
- **Experimental Materials Science** [Christof Wöll (KIT), Christoph Koch (HU Berlin), Claudia Felser (MPI-CPfS), Dierk Raabe (MPI-E), et al.]: Electron microscopy, atom-probe, AFM, various spectroscopic methods (incl. synchrotrons), etc.
- **Soft-Matter Research** [Kurt Kremer und Tristan Bereau (MPI for Polymer Research), Carsten Baldauf (U Leipzig und FHI Berlin), et al.]: Polymers, gels, liquids, biomolecules, cells, tissues, etc.
- **Catalysis** [Karsten Reuter (TU Munich), Roger Gläser (U. Leipzig), Frank Rosowski (BasCat und TU Berlin), Robert Schlögl (MPI-CEC), Stephan Schunk (hte/BASF) et al.]: Kinetic and spectroscopic experiments, structure and composition studies, multiscale modeling, etc.
- **Digital Infrastructure** [Hans Bungartz (TU Munich), Wolfgang Nagel (TU und ZIH Dresden)]

Kolleginnen und Kollegen sind eingeladen, ihre Anforderungen zu beschreiben und ihre Expertise einzubringen, um die nötigen Konzepte und Dateninfrastrukturen aktiv mitzugestalten.

Kontaktieren Sie uns dazu unter fairmat@fairdi.eu oder bekunden Sie Ihr Interesse über unsere Webpage: <https://fairdi.eu/fairmat>.

Folgende Institutionen haben bislang ihre Unterstützung bzw. Mitarbeit in FAIRmat zugesagt:

Fritz-Haber-Institut der Max-Planck-Gesellschaft 	Fritz-Haber-Institut der Max-Planck-Gesellschaft	Max-Planck-Institut für Polymerforschung 	MAX-PLANCK-INSTITUT FÜR POLYMERFORSCHUNG MAX PLANCK INSTITUTE FOR POLYMER RESEARCH
Max Planck Computing & Data Facility (MPCDF) 	MPCDF	Max-Planck-Institut für Chemische Physik fester Stoffe 	MAX-PLANCK-INSTITUT FÜR CHEMISCHE PHYSIK FESTER STOFFE
Max-Planck-Gesellschaft, Netzwerk zu <i>Big-Data-Driven Materials Science</i> 	BIG max MAX PLANCK RESEARCH NETWORK on big-data-driven materials science	Humboldt-Universität zu Berlin 	HUMBOLDT-UNIVERSITÄT ZU BERLIN
Technische Universität München 	TUM	Karlsruher Institut für Technologie 	KIT Karlsruher Institut für Technologie
BasCat – UniCat BASF Joint Lab, Technische Universität Berlin 	BasCat UniCat BASF JointLab	Max-Planck-Institut für Plasmaphysik (IPP) 	IPP
Paul-Drude-Institut für Festkörperelektronik 	pfi	Deutsche Gesellschaft für Katalyse 	GE C A T S GERMAN CATALYSIS SOCIETY
Dresden Center for Computational Materials Science 	DCMS	Universität Leipzig 	UNIVERSITÄT LEIPZIG
Friedrich-Schiller-Universität Jena 	FRIEDRICH-SCHILLER-UNIVERSITÄT JENA	Max-Planck-Institut für Festkörperforschung 	Max-Planck-Institut für Festkörperforschung
FDMentor 	FDMentor	Leibniz-Institut für Kristallzüchtung 	IKZ

<p>Helmholtz Zentrum Berlin</p> 	<p>Max Planck Institut für Kolloid- und Grenzflächenforschung</p> 
<p>Bundesanstalt für Materialforschung und -prüfung</p> 	<p>Zuse-Institut Berlin</p> 