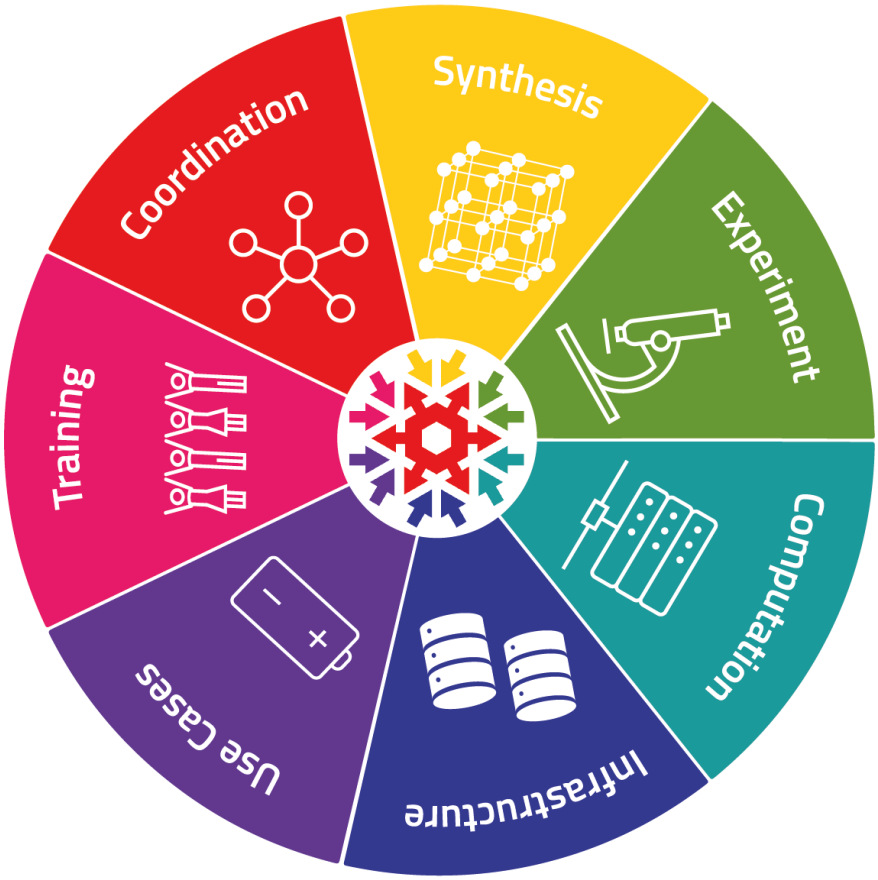
FAIR Data Infrastructure for Condensed-Matter Physics, and the Chemical Physics of Solids

FAIRmat Progress Report



FAIR-DI

Karlsruher Institut für Technologie

B-1 FAIRmat Progress Report - Part 1

Table of Contents

1 General Information

- Name of the consortium

FAIRmat – FAIR Data Infrastructure for Condensed-Matter Physics, and the Chemical Physics of Solids

- Research domains or research methods addressed by the consortium

FAIRmat builds an infrastructure for Findable, Accessible, Interoperable, and Reusable (FAIR) data in **condensed-matter physics** (DFG Review Board 307) and **the chemical physics of solids** (DFG Review Board 304). It establishes a research data management (RDM) service for collecting, organizing, sharing, analyzing, and publishing FAIR data across **synthesis**, **experiment**, **theory**, and **computation**.[1] We demonstrate the use of RDM on **batteries**, **solar cells**, **catalysis**, **metal-organic frameworks**, and **biophysical materials**. We also enable progress in materials informatics through the use and development of Artificial Intelligence methods. Due to conceptual overlaps between different domains and the generality of our approach, FAIRmat inherently impacts other physical sciences and interdisciplinary sub-domains, e.g., materials science and chemistry, thus benefiting engineering, industry, and society.

- URL of the consortium website and repositories used for publishing output

Name	Description	Link
FAIRmat	Website of the consortium	https://www.fairmat-nfdi.eu/fairmat
NOMAD	NOMAD's services website	https://nomad-lab.eu/nomad-lab
NOMAD	Archive and Repository, dedicated to publishing and exploring data	https://nomad-lab.eu/prod/v1/gui/search/entries
FAIRmat GitHub organization	A collection of open-source code repositories from FAIRmat	https://github.com/FAIRmat-NFDI
FAIRmat community on Zenodo	A collection of presentations, posters, guides, newsletters, and other outputs from FAIRmat members and the community	https://zenodo.org/communities/fairmat_nfdi

[1] M. Scheffler, M. Aeschlimann, M. Albrecht, T. Bereau, H.-J. Bungartz, C.Felser, M. Greiner, A. Groß, C. Koch, K. Kremer, W. E. Nagel, M. Scheidgen, C. Wöll, and C. Draxl, FAIR data enabling new horizons for materials research, Nature 604, 635 (2022). https://doi.org/10.1038/s41586-022-04501-x

2 Summary

FAIRmat started with the goal of supporting condensed-matter physics and the chemical physics of solids with **a single integrated**, **extensible**, **distributed FAIR RDM service**. This vision has been implemented in our data infrastructure NOMAD.[2] What began as a repository for publishing *ab initio* calculations,[3] has been turned into a comprehensive RDM service covering (1) the definition of data schemas, (2) the data acquisition from a wide range of sources, (3) the processing of files into structured and reusable data, (4) the sharing of data across research institutions, and (5) the analysis of data. FAIRmat enables the community to build a network of interconnected NOMAD instances, called Oases, that balance the use of local resources, customizable services, and data privacy on the one hand with easy collaboration, ready-made tools, and data sharing on the other. **Figure 1** shows key components and figures.

NOMAD is built on **structured data based on well-defined schemas**. It not only captures shallow user-provided metadata, such as authors, dates, or keywords, but also automatically extracts deep metadata, including parameters, settings, measurements, logs, workflows, and analysis results. NOMAD extracts more than 8,100 different quantities, with new plugins being developed continuously. This approach has allowed us to provide support for over **141 file formats** of specific instruments and simulation codes, add data from **multiple external databases** (Materials Project, Aflow, OQMD, EELS, and HTEM), support community-relevant **data standards** (OPTIMADE, NeXus, .eln, …), and offer electronic lab notebook (**ELN**) **data import** (eLabFTW, Chemotion, LabFolder, and openBIS). Since all data are processed and uniformly structured, we can also offer a comprehensive suite of **visualization**, **machine learning**, **and Artificial Intelligence tools**. All tools are containerized and run directly in any browser.

Since research is highly diverse and specialized, all schemas, data, and tools need to be highly specialized as well. Therefore, FAIRmat selected relevant research and application subdomains and created corresponding **NOMAD plugins** and **datasets**. These demonstrate the impact of our FAIR and structured approach and convince communities to invest in and adapt our solutions for other domains. For example, we have built the NOMAD perovskite database, which contains 43,000 solar cells, originally curated from “un-FAIR” data published in thousands of papers. Now, all solar cells can be explored from a single interface with hundreds of filters and

[2] M. Scheidgen, L. Himanen, A. N. Ladines, D. Sikter, M. Nakhaee, Á. Fekete, T. Chang, A. Golparvar, J. A. Márquez, S. Brockhauser, S. Brückner, L. M. Ghiringhelli, F. Dietrich, D. Lehmberg, T. Denell, A. Albino, H. Näsström, S. Shabih, F. Dobener, M. Kühbach, R. Mozumder, J. F. Rudzinski, N. Daelman, J. M. Pizarro, M. Kuban, C. Salazar, P. Ondračka, H.-J. Bungartz, and C. Draxl, NOMAD: A distributed web-based platform for managing materials science research data, J. Open Source Softw. 8, 5388 (2023). https://doi.org/10.21105/joss.05388

[3] C. Draxl and M. Scheffler, The NOMAD Laboratory: From Data Sharing to Artificial Intelligence, J. Phys. Mater. 2, 036001 (2019). https://doi.org/10.1088/2515-7639/ab13bb

interactive statistics. All metadata are consistent, and researchers can even add newly reported solar cells. Other notable examples include a database of 13 million DFT calculations; a database of 17,000 metal-organic frameworks (MOFs); a framework for managing and exploring data from heterogeneous catalysis; an atom probe tomography dataset curated from files published on Zenodo and standardized in the NeXus format; plugins supporting synthesis workflows from leading national and international laboratories, including the Leibniz-Institut für Kristallzüchtung, the Helmholtz-Zentrum Berlin, and the Technical University of Denmark; measurements from popular instruments from vendors such as Bruker, Malvern Panalytical, or Thermo Fisher Scientific; and our integrated lab automation software, NOMAD CAMELS. All of these examples typically come with dedicated schemas, parsers, and normalizers for processing the respective files, visualization, analysis tools, and published datasets.

FAIRmat invests heavily in training the scientific community to appreciate the value of FAIR data by providing comprehensive online tutorials and user training, reaching out to academia and companies at conferences and dedicated workshops, and by deploying specially configured NOMAD Oasis installations to enable digitalization of practical courses at universities.

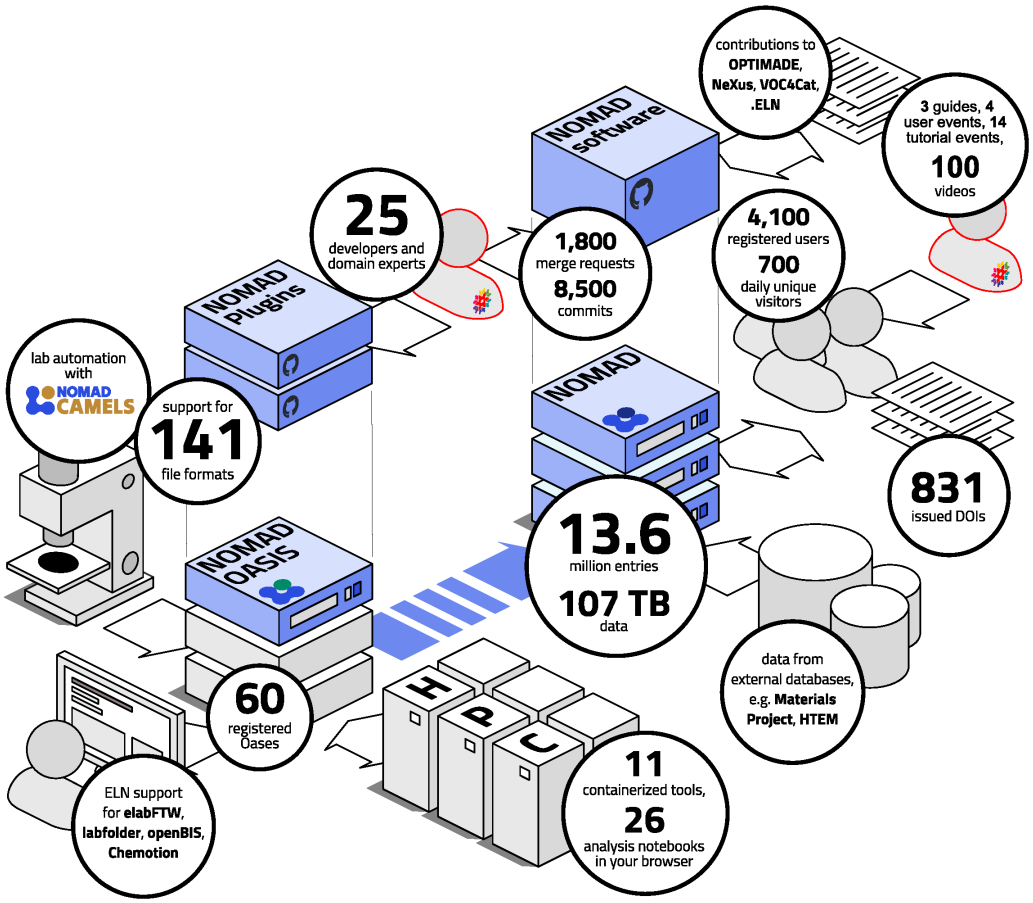


Figure 1: The integrated, extendable, and distributed RDM service NOMAD developed by FAIRmat.

3 Composition of the Consortium

- Applicant institution

Applicant institution	Location	Duration
Humboldt-Universität zu Berlin (HU Berlin)	Berlin	10/21-09/26

- Spokesperson

Spokesperson	Institution, location	Duration	ORCID-ID
Claudia Draxl	HU Berlin, Berlin & FAIR-DI e.V., Berlin	10/21-09/26	https://orcid.org/0000-0003-3523-6657

- Co-applicant institutions

Co-applicant institutions	Location	Duration
Leibniz-Institut für Kristallzüchtung (IKZ)	Berlin	10/21-09/26
Max Planck Institute for Chemical Energy Conversion (MPI-CEC)	Mühlheim/Ruhr	10/21-04/22
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)	Erlangen-Nürnberg	08/22-09/26
Fritz Haber Institute of the Max Planck Society (FHI)	Berlin	10/21-03/22
Ruhr-University Bochum (RUB)	Bochum	03/22-09/26
Technical University of Munich (TUM)	Garching	10/21-09/26
Karlsruhe Institute of Technology (KIT)	Karlsruhe	10/21-09/26
FAIR-DI e.V.	Berlin	10/21-09/26

- Co-spokespersons

Co-spokespersons	Institution, location	Task area	Duration	ORCID-ID
Martin Albrecht	IKZ, Berlin	Area A	10/21-09/26	https://orcid.org/0000-0003-1835-052X
Mark Greiner	MPI-CEC, Mülheim/Ruhr	Area B	10/21-04/22	https://orcid.org/0000-0002-4363-7189
Heiko Weber	FAU, Erlangen	Area B	08/22-09/26	https://orcid.org/0000-0002-6403-9022
Matthias Scheffler	FHI, Berlin	Area C & F	10/21-03/22	https://orcid.org/0000-0002-1280-9873
Silvana Botti	RUB, Bochum	Area C	03/22-09/26	https://orcid.org/0000-0002-4920-2370
Hans-Joachim Bungartz	TUM, Garching	Area D	10/21-09/26	https://orcid.org/0000-0002-0171-0712
Christof Wöll	KIT, Karlsruhe	Area E	10/21-09/26	https://orcid.org/0000-0003-1078-3304

- Participants

Participating institutions	Location	Duration
German Physical Society (DPG)	Bad Honnef	10/21-09/26
Leibniz Information Centre for Science and Technology (TIB)	Hannover	10/21-09/26
Max Planck Computing and Data Facility (MPCDF)	Garching	10/23-09/26

Participating individuals	Institution, location	Duration	ORCID-ID
Martin Aeschlimann	University of Kaiserslautern-Landau, Kaiserslautern	10/21-09/26	https://orcid.org/0000-0003-3413-5029
Sören Auer	TIB, Hannover	10/21-09/26	https://orcid.org/0000-0002-0698-2864
Carsten Baldauf	FHI, Berlin	10/21-09/26	https://orcid.org/0000-0003-2637-6009
Tristan Bereau	Heidelberg University, Heidelberg	10/21-01/22 06/23-09/26	https://orcid.org/0000-0001-9945-1271
Stefan Blügel	Research Center Jülich (FZJ), Jülich	10/21-09/26	https://orcid.org/0000-0001-9987-4733
Silvana Botti	RUB, Bochum	10/21-09/26	https://orcid.org/0000-0002-4920-2370
Christoph J. Brabec	FAU, Erlangen	10/21-09/26	https://orcid.org/0000-0002-9440-0253
Victoria Coors	HU Berlin, Berlin	10/21-09/26	/
Malte Dreyer	HU Berlin, Berlin	10/21-09/26	https://orcid.org/0000-0002-1775-8622
Natascha Dropka	IKZ, Berlin	10/21-09/26	https://orcid.org/0000-0002-1530-7449
Ralph Ernstorfer	FHI, Berlin	10/21-11/22	https://orcid.org/0000-0001-6665-3520
Norbert Esser	Leibniz Institute for Analytical Sciences, Berlin	10/21-07/23	https://orcid.org/0000-0003-0345-6823
Claudia Felser	Max Planck Institute for Chemical Physics of Solids, Dresden	10/21-09/26	https://orcid.org/0000-0002-8200-2063
Luca Ghiringhelli	FAU, Erlangen	10/21-09/26	https://orcid.org/0000-0001-5099-3029
Roger Gläser	Leipzig University (U Leipzig), Leipzig	10/21-09/26	https://orcid.org/0000-0002-8134-4280
Axel Groß	Ulm University, Ulm	10/21-09/26	https://orcid.org/0000-0003-4037-7331
Marius Grundmann	U Leipzig, Leipzig	10/21-09/26	https://orcid.org/0000-0001-7554-182X
Aleksander Gurlo	Technical University Berlin, Berlin	10/21-09/26	https://orcid.org/0000-0001-7047-666X
Thomas Hammerschmidt	RUB, Bochum	10/21-09/26	https://orcid.org/0000-0002-2270-4469
Tamás Haraszti	Leibniz Institute for Interactive Materials (DWI), Aachen	10/21-09/26	https://orcid.org/0000-0002-7095-4358
Stefan Hecht	HU Berlin, Berlin	10/21-09/26	https://orcid.org/0000-0002-6124-0222
Thomas Heine	Technical University Dresden (TU Dresden), Dresden	10/21-09/26	https://orcid.org/0000-0002-7109-2788
Kerstin Helbig	HU Berlin, Berlin	10/21-09/26	https://orcid.org/0000-0002-2775-6751
Walid Hetaba	MPI-CEC, Mühlheim/Ruhr	07/22-09/26	https://orcid.org/0000-0003-4728-0786
Kevin Jablonka	Friedrich Schiller University Jena, Jena	11/23-09/26	https://orcid.org/0000-0003-4894-4660

Jürgen Janek	Justus Liebig University Giessen, Giessen	10/21-09/26	https://orcid.org/0000-0002-9221-4756
Heinz Junkes	FHI, Berlin	10/21-09/26	https://orcid.org/0000-0002-0218-4873
Josef A. Käs	U Leipzig, Leipzig	10/21-11/23	https://orcid.org/0000-0003-3158-2480
Christoph T. Koch	HU Berlin, Berlin	10/21-09/26	https://orcid.org/0000-0002-3984-1523
Sarah Köster	Georg-August-University Göttingen, Göttingen	10/21-09/26	https://orcid.org/0000-0002-0009-1024
Kurt Kremer	Max Planck Institute for Polymer Research, Mainz	10/21-09/26	https://orcid.org/0000-0003-1842-9369
Michael Krieger	FAU, Erlangen	10/21-09/26	https://orcid.org/0000-0003-1480-9161
Markus Kühbach	HU Berlin, Berlin	10/21-09/26	https://orcid.org/0000-0002-7117-5196
Erwin Laure	MPCDF, Garching	02/22-09/26	https://orcid.org/0000-0002-9901-9857
Hermann Lederer	MPCDF, Garching	10/21-02/22	https://orcid.org/0000-0003-1978-2703
Miguel A. L. Marques	RUB, Bochum	04/22-09/26	https://orcid.org/0000-0003-0170-8222
Ingrid Mertig	Martin-Luther-University Halle-Wittenberg, Halle	10/21-09/26	https://orcid.org/0000-0001-8488-0997
Wolfgang E. Nagel	TU Dresden, Dresden	10/21-09/26	/
Rossitza Pentcheva	University Duisburg-Essen, Duisburg	10/21-09/26	https://orcid.org/0000-0002-4423-8980
Dierk Raabe	Max-Planck-Institut für Eisenforschung GmbH, Düsseldorf	10/21-09/26	https://orcid.org/0000-0003-0194-6124
Alexander Reinefeld	Zuse Institute Berlin, Berlin	10/21-02/22	/
Laurenz Rettig	FHI, Berlin	02/23-09/26	https://orcid.org/0000-0002-0725-6696
Karsten Reuter	FHI, Berlin	10/21-09/26	https://orcid.org/0000-0001-8473-8659
Raphael Ritz	MPCDF, Garching	10/21-09/26	https://orcid.org/0000-0003-4615-6804
Cesar Rodriguez-Emmenegger	DWI, Aachen	10/21-09/26	https://orcid.org/0000-0003-0745-0840
Erich Runge	ex officio, DPG, Bonn	10/21-09/26	https://orcid.org/0000-0002-2157-229X
Markus Scheidgen	HU Berlin, Berlin	10/21-09/26	https://orcid.org/0000-0002-8038-2277
Robert Schlögl	FHI, Berlin	10/21-02/24	https://orcid.org/0000-0002-5163-1051
Thomas Schröder	IKZ, Berlin	10/21-09/26	https://orcid.org/0000-0002-7547-3828
Erdmann Spiecker	FAU, Erlangen	05/23-09/26	https://orcid.org/0000-0002-2723-5227
Godehard Sutmann	FZJ, Jülich	10/21-09/26	https://orcid.org/0000-0002-9004-604X

Huayna Terraschke	Christian-Albrechts-Universität zu Kiel, Kiel	11/23-09/26	https://orcid.org/0000-0002-5479-7663
Annette Trunschke	FHI, Berlin	10/21-09/26	https://orcid.org/0000-0003-2869-0181
Thomas Unold	Helmholtz Zentrum Berlin, Berlin	10/21-09/26	https://orcid.org/0000-0002-5750-0693
Denis Usvyat	HU Berlin, Berlin	05/24-09/26	https://orcid.org/0000-0001-6351-9289
Roser Valenti	Goethe University Frankfurt, Frankfurt/Main	10/21-09/26	https://orcid.org/0000-0003-0497-1165
Holger von Wenck-stern	U Leipzig, Leipzig	10/21-09/26	https://orcid.org/0000-0002-3936-275X
Dieter Weber	FZJ, Jülich	10/21-03/23	https://orcid.org/0000-0001-6635-9567
Heiko Weber	FAU, Erlangen	10/21-09/26	https://orcid.org/0000-0002-6403-9022
Stefan Wesner	University of Cologne, Köln	10/21-09/26	https://orcid.org/0000-0002-7270-7959
Joachim Wosnitza	Helmholtz Zentrum Dresden-Rossendorf, Dresden	10/21-09/26	https://orcid.org/0009-0008-4136-5643
Hongbin Zhang	Technical University of Darm-stadt	08/23-09/26	https://orcid.org/0000-0002-1035-8861