



FAIRmat Newsletter

VOLUME 4 | DECEMBER 2023

Editorial



Welcome to the latest edition of the FAIRmat Newsletter, which serves as a showcase for the diversity of research data management (RDM) within the FAIRmat initiative. This issue provides a summary of the profound progress that has been made over the past six months. In addition to several recurring segments in this issue, we introduce you to new faces joining the FAIRmat leadership, report on our participation in the 1st Conference on Research Data Infrastructure (CoRDI), and give an overview of our third FAIRmat users meeting and the CE-CAM flagship workshop: "FAIR and TRUE Data Processing for Soft Matter Simulations". We inform you about our first publication in the Journal of Open Source Software (JOSS) and update you on recent collaborations with other consortia of the National Research Data Infrastructure (NFDI).

We encourage you to use the articles in this newsletter as inspiration for your own research and to actively participate in the discussion and exchange within the FAIRmat community.

We thank all the authors who have contributed to this issue and look forward to more inspiring developments and discoveries as we move towards ever more advanced FAIR research data management in materials science.

Martin Aeschlimann, leader of Area F



FAIRmat news

FAIRmat participation at CoRDI 2023

From September 12-14, 2023, the association National Research Data Infrastructure (NFDI) hosted the 1st Conference on Research Data Infrastructure (CoRDI) at the Karlsruhe Institute of Technology. FAIRmat was well represented with talks and posters. Heiko Weber (leader of Area B: Experiment) presented "[Research Data Management for Experiments in Solid-State Physics: Concept](#)", and Markus Scheidgen (infrastructure coordinator of Area D: Infrastructure) introduced "[FAIR Research Data with NOMAD](#)". Poster presentations included "[FAIRmat Guide to Writing Data Management Plans](#)" by Ahmed Mansour (coordinator of Area F: Training) and "[Towards FAIR Data in Heterogeneous Catalysis Research](#)" by Julia Schumann (catalysis expert in Area E: Use Cases). Beyond formal presentations, our outreach team engaged with participants at the FAIRmat stand in the "Marketplace of the Consortia", discussing FAIRmat and NOMAD and exploring potential collaboration opportunities.



New faces join FAIRmat leadership

FAIRmat has recently welcomed three new task leaders. Huayna Terraschke joined [Area A](#) as a leader for task A3: Synthesis from solid phase and solution, Tristan Berau joined [Area C](#) as a leader for task C2: Classical simulations and multi-scale modeling, and Kevin Jablonka joined [Area E](#) as a leader for task E7: Artificial-intelligence toolkit.

In addition, the position of the scientific coordinator was filled by José Márquez Prieto. He had worked as a domain expert in the field of optoelectronic materials.

We welcome all new members to the team!

Project milestones

NOMAD is published in the Journal of Open Source Software

Our web-based software for managing materials data, NOMAD, has been published in the Journal of Open Source Software (JOSS). JOSS is an open-access and peer-reviewed academic journal that publishes research software packages crucial to scientific research. JOSS has an open review process which takes place in GitHub. [Link to the publication.](#)

Extended support for computational workflows in NOMAD

Traditionally, a repository for contributions from density-functional theory calculations, NOMAD's computational support has been extensively broadened. NOMAD includes, for instance, the integration of advanced many-body calculations (GW, BSE, DMFT), classical molecular-dynamics simulations, and complex simulation workflows. These new developments are already being utilized, as demonstrated by the recent massive upload of the [molecular dynamics dataset](#) "Atomistic Molecular Dynamics Simulations of Pure Liquids and Binary Mixtures for Representative C₇O₂ Isomers", which was recently used by the group of Tristan Bereau to parameterize a transferable coarse-grained model and contains 4,168 entries. This dataset takes advantage of the custom workflow functionality in NOMAD, and represents the potential utility of NOMAD within a broad range of computational sub-domains.

Interactive Metal-Organic Framework (MOF) search dashboard

A [search app for MOFs](#) was deployed as the second interactive dashboard use case in NOMAD. It is now available in its beta version. Customization with interactive widgets allows users to search, navigate, and visualize MOF data related to various properties. The app currently contains numerous calculated MOFs and will soon be expanded to include experimental synthesis conditions. The app is based on a NOMAD normalizer that recognizes MOFs and other porous materials such as zeolites and covalent organic frameworks, removes unbound guest molecules, decomposes MOFs into their constituent units, and calculates geometric properties of MOFs.

The FAIRmat infrastructure



Markus Scheidgen
infrastructure coordinator

Customizable search and visualization

FAIRmat builds an extensible and customizable data infrastructure. Following a bottom-up approach, we develop the NOMAD research data management software independent of specific schemas, file formats, and workflows. Users, data stewards, and administrators can add support for their own data. They develop schemas, parsers, and plugins to adapt their NOMAD for their needs. Adding custom schemas and parsers provides excellent utility for adding new data types, especially for experiments and synthesis. But how can we explore new types of data? How can we build a detailed faceted search interface, if the specific data structure is not fixed? How can we provide data-type specific visualization if we need to know all data types in advance?

With the latest release of NOMAD, we added custom search applications for data based on uploaded and plugin schemas. This custom search application allows users to search for scalar quantities. More importantly, it will enable data stewards and NOMAD administrators to configure specialized search applications that can include custom quantities. Previously, only specific quantities from core NOMAD schemas could be used. Now, any data can be used to create search filters and build dashboards from histograms and other visualizations to explore datasets. Another recently added feature is the redesigned support for custom plots. Schemas can be augmented with annotations that describe plots based on the defined data. While previously, only simple line plots could be added, now the full potential of the powerful and popular [plotly](#) library can be used.



Meet our users



Daniel Baumann

Doctoral researcher in the group Next Generation Photovoltaics at Karlsruhe Institute of Technology

What is the focus of your doctoral research?

I am focusing on advancing perovskite solar-cell technology through robotic spin-coating. My primary objective is to reduce human error while enhancing repeatability and reproducibility in manufacturing. I plan to incorporate learning algorithms to speed up materials research within the perovskite solar-cell field. This approach not only streamlines processes but also sets the stage for optimization. By integrating automation, the project aims at improving the efficiency and reliability of perovskite solar-cells.

What challenges do you think doctoral researchers face when applying the FAIR principles to their data?

The absence of established infrastructure presents a significant challenge for individual researchers in following FAIR principles for data management. Researchers often need too much time to organize their data in a FAIR manner, especially when dealing with diverse tools and formats from measurement systems. A predefined group infrastructure makes it realistic for individuals to store and manage data in a FAIR manner. The overarching problem lies in the absence of a standardized structure; each researcher develops their system, resulting in non-compatibility and a lack of motivation to adhere to FAIR principles unless there is a direct benefit. Overcoming these challenges requires establishing user-friendly infrastructures like NOMAD that demonstrate individual benefits, emphasizing usability over the sole purpose of publishing FAIR data for knowledge advancement.

What is the strategy for RDM in your group and how is FAIRmat helping in implementing it?

Previously, during my master's thesis, my research data management was done manually, with each experiment individually evaluated. The time-consuming manual process became less appealing with the move towards automation. I had to decide whether to handle everything in Python on my own or leverage existing structures. My research data management during my doctoral research involves implementing a NOMAD Oasis for planning, storing, and evaluating data. After a positive experience, more colleagues will soon join and use NOMAD Oasis. FAIRmat and Michael Götte from HZB have assisted in setting up the NOMAD Oasis, providing customizations and extensive bug fixing, which was a crucial step for my highly positive user experience and my supervisors' decision to

expand the usage of NOMAD to the whole research group.

Meet our experts



Adrianna Wojas

UX/UI expert in Area D

What is it that you do here at FAIRmat?

At FAIRmat, I work on improving the user experience for the NOMAD application and various project websites. My role focuses on working closely with the development team and researchers to ensure that we present data and research in a user-friendly and visually appealing way. As I am passionate about psychological patterns and accessibility indicators, I am constantly looking for ways to integrate them into our software.

My daily tasks can range from discussions with researchers to wireframing, optimizing documentation, and asking an enormous amount of questions to continually improve our user experience. It is a dynamic role that I genuinely enjoy.

What drew you to join FAIRmat?

Before joining FAIRmat, I worked as a UX/UI specialist in the tech industry, where I honed my skills in designing interfaces for AI-driven applications and exploring the exciting field of augmented reality. While I enjoyed my previous role, I was eager to apply my skills to more meaningful and inspiring projects.

With a background in fine art and a strong passion for visual things, I have always been a pragmatic person seeking to understand how things work and improve them. I am particularly drawn to projects that prioritize accessibility and share a commitment to the values of knowledge and education. When I learned about FAIRmat's commitment to making research data more accessible and understandable through the FAIR principles, I was honestly intrigued. It looked like a fantastic project that had the potential to make a real and positive impact, so I applied.

What is your favorite thing about working at FAIRmat?

What I appreciate most about FAIRmat is the constant encouragement to explore, learn, and grow. My role has allowed me to dive into the fascinating world of data visualization, a field I have come to love. It has also challenged me with data analysis and graphing, which is sometimes tricky but incredibly satisfying.

What makes this journey even better is the fantastic team I work with. I am lucky to have colleagues who support experimentation, whether it is in design or coding aspirations. I never hesitate to ask for help or explanations - everyone is happy and enthusiastic to help and explain (even if I ask many questions!).

To sum up, it is a supportive and motivating atmosphere at FAIRmat. We are an incredible team of open-minded and inspiring people who respect and encourage each other to progress professionally and personally.

Opinion article



Eva Unger

Group leader "Solution-Processing of Hybrid Materials and Devices" at Helmholtz-Zentrum Berlin, professor at the Institute for Chemistry, Humboldt-Universität zu Berlin, and lecturer at Lund University (Sweden).

Verification by FAIRification of PV research data management

For years, a collaborative effort has flourished among principal investigators across various Helmholtz Centers in Germany, aiming to enhance collaboration and impact in photovoltaic (PV) research. The Helmholtz Center Berlin (HZB), Forschungszentrum Jülich (HI-ERN), and the Karlsruhe Institute of Technology (KIT) have jointly launched projects such as "Emerging PV" and "The Perovskite Database", creating secondary dissemination platforms for research data. The ultimate vision is to establish collaborative research-data-sharing platforms tailored to the unique needs of the PV research community. We are further exploiting these efforts to export the use of PV-specific NOMAD data infrastructure to other research organizations to support global initiatives in the FAIRification of research data, such as within the VIPERLAB EU project.

The FAIRmat project has emerged as a strategic partner in this mission, providing invaluable support to meet specific research data management needs. Through smart recruitment decisions, the collaboration is now implementing NOMAD, establishing NOMAD Oasis in key research infrastructures like the HySPRINT laboratory at HZB. Simultaneously, efforts are underway to align on a common data ontology for perovskite PV, compare Electronic Lab Notebooks (ELNs), and interchange software tools. This alignment facilitates a standardized approach, enhancing the reproducibility and verifiability of research data.

NOMAD, a robust research data management soft-

ware, captures metadata related to sample fabrication and measurements. By emulating modern ELNs and combining various functionalities, NOMAD streamlines data recording and promotes collaboration among team members. The platform serves as a centralized hub for managing experimental information, demonstrating its potential as a comprehensive research data management ecosystem.

The ongoing effort includes implementing NOMAD Oasis in various PV research infrastructures at our respective institutions, ensuring seamless integration into existing workflows. We found that engaging data stewards in the process of implementing the research data platforms has played a crucial role: the customization of NOMAD to suit specific experimental workflows within research organizations requires significant adaptation. Training initiatives and incitement efforts are essential to motivate experimental scientists to use NOMAD for their research data management. We have put much focus in reducing hurdles to enable users to have positive experiences in NOMAD, making their daily laboratory work more effective.

In conclusion, the collaborative initiatives among PV researchers of the Helmholtz Association of Germany have made a significant leap forward by utilizing NOMAD and the support of FAIRmat. Key activities included defining a standardized data ontology, starting local initiatives using and setting up a NOMAD Oasis reflecting our laboratories' specific workflows, aligning and interchanging ELNs, and sharing software tools. We hope that this will soon translate to an enhanced reproducibility and verifiability of our research data and, as a result, will increase the impact and importance of research data generated within our institutions, contributing to a general culture of open and efficient scientific research in the global PV community.



Visualization of the FAIRmat family for a FAIR data infrastructure - NOMAD, NOMAD Oasis, and NOMAD CAMELS by Adrianna Wojas.

What is: a PID?



Joseph Rudzinski
Area C coordinator

A Persistent Identifier (PID) is a unique, long-lasting label assigned to a digital object. PIDs serve as digital fingerprints and permanent addresses, ensuring digital objects remain identifiable and findable, even as websites or data repositories change over time. In the scientific realm, PIDs are often given to datasets, publications, or research resources in the form of Digital Object Identifiers (DOIs). PIDs can uniquely identify individual researchers, e.g., Open Researcher and Contributor IDs (ORCID IDs). They enable reliable and consistent identification of scientific data, and make sharing, citing, and tracking the usage and impact of data and publications more straightforward. Thus, PIDs are vital for the reproducibility and credibility of scientific findings and are an essential to FAIR data management.

Collaborations with other NFDI consortia

Physical Sciences in NFDI

In the dynamic collaboration, Physical Sciences in NFDI ([PSinNFDI](#)), we have united with fellow NFDI consortia, including [DAPHNE4NFDI](#), [MaRDI](#), [NFDI-MatWerk](#), [NFDI4Cat](#), [NFDI4Chem](#), and [PUNCH4NFDI](#). This overarching initiative promotes collaboration among the physical sciences in the [NFDI](#). This autumn, PSinNFDI hosted its first webinar entitled “Unlocking the Potential of Data – FAIR Data Principles in NFDI”, drawing over 50 participants and engaging in enriching discussions. FAIRmat took center stage, featuring contributions by José Márquez Prieto on NOMAD’s Solar Cell App and Alexander Fuchs on NOMAD CAMELS. The recorded webinar is now accessible [online](#).



Outreach and training

Third FAIRmat users meeting

The third FAIRmat users meeting took place November 15-16, 2023, in Berlin, attracting a diverse group of attendees, including NOMAD users, collaborators, researchers, representatives from joint projects supporting FAIRmat, and various NFDI consortia. This two-day public event provided a unique platform for community building, training, and networking, fostering the exchange of valuable insights into research data management and the NOMAD infrastructure.



The program started with invited talks, spanning perspectives from early career researchers to policymakers. Michael Mößle, DFG program director, outlined the current status and future perspectives of the NFDI, offering a funder’s perspective into research data management. José Márquez Prieto, FAIRmat’s scientific coordinator, provided an overview of FAIRmat and NOMAD. NOMAD users, Pavel Ondračka and Daniel Baumann, shared inspiring presentations on how NOMAD enhances their research activities. The talks were streamed online and will be soon to be available on our [YouTube channel](#).

The first day concluded with a dynamic poster session including live demonstrations of the latest NOMAD features and NOMAD CAMELS. The second day featured interactive workshops where participants engaged in hands-on activities with NOMAD and exploring research data management practices.

CECAM flagship workshop: “FAIR and TRUE Data Processing for Soft Matter Simulations”

The CECAM flagship workshop, co-organized by FAIRmat and the [MoSDeF group](#), hosted by the Max Planck Institute for Polymer Research, recently drew a diverse group of people within the molecular dynamics (MD) community together. This pioneering event focused on addressing challenges in the *FAIRification* of molecular simulations.

Structured as a discussion-focused initiative, the workshop featured invited talks, Q&A sessions, and five round-table discussions. Topics encompassed data prov-

enance strategies, simulation (meta)data storage, simulation engine interoperability, and data structures for edge cases. Tristan Bereau (task leader in Area C), showcased his group's groundbreaking dataset of C_7O_2 isomers, underlining NOMAD's potential as a valuable tool for the MD community.



The workshop concluded optimistically, paving the way for future collaborations between software projects. FAIRmat eagerly anticipates contributing to the ongoing advancements in FAIR-data management within the MD community.

Seminar series

FAIRmat has recently launched a new outreach and communication format - the FAIRmat Seminar series. We regularly invite researchers from different fields within natural science and digital infrastructures to Berlin to give a talk. In addition, the seminar talks are streamed online and the recordings are available on our [YouTube channel](#).

The inaugural seminar on "High throughput search of topological material and meta-materials" was given by Maria Vergniory in July 2023. In September, two seminars followed: Kevin Jablonka took us into digital chemistry and large language models with "Why machine learning can find a new material but not a needle in a haystack". Brian Richard Pauw talked about "Glimpses of the future: A full-stack, highly automated materials research laboratory".

We have already scored a hat trick and are looking forward to the next seminar episodes in 2024.

Upcoming events

DPG Spring meeting in Berlin

The DPG spring meeting of the Condensed Matter Section (SKM) will take place at the Technische Universität Berlin, March 17-23, 2024. Besides individual contributions by FAIRmat members, we will give a tutorial and share a booth with DAPHNE4NFDI. Visit us there for live demonstrations and advice on our tools! [Find out more!](#)

International FAIR Digital Objects Implementation Conference

The International FAIR Digital Objects Implementation Conference will take place from March 19-21, 2024 in Berlin. The FDO2024 conference, unites science, industry, and policymakers to advance FAIR Digital Objects (FDOs). This edition focuses on practical FDO implementations, addressing challenges in creating an interoperable, stable, and persistent global digital data domain. The key events will unfold at the Museum für Naturkunde Berlin and the DIN standardization organization, fostering collaboration and innovation in digital data. [Find out more!](#)

Recent publications

- ❄ L.M. Ghiringhelli, C. Baldauf, T. Bereau, S. Brockhauser, C. Carbogno, J. Chamanara, S. Cozzini, S. Curtarolo, C. Draxl, S. Dwaraknath, Á. Fekete, J. Kermode, C.T. Koch, M. Kühbach, A.N. Ladines, P. Lambrix, M.O. Lenz-Himmer, S. Levchenko, M. Oliveira, A. Michalchuk, R. Miller, B. Onat, P. Pavone, G. Pizzi, B. Regler, G.M. Rignanese, J. Schaarschmidt, M. Scheidgen, A. Schneidewind, T. Sheveleva, C. Su, D. Usvyat, O. Valsson, C. Wöll, and M. Scheffler, [Shared Metadata for Data-Centric Materials Science](#), *Sci. Data* 10, 626 (2023).
- ❄ M. Jalali, A.D. D. Wonanke, C. Wöll, [MOFGalaxyNet: a social network analysis for predicting guest accessibility in metal-organic frameworks utilizing graph convolutional networks](#), *J. Cheminform.* 15, 94 (2023).
- ❄ M. Scheidgen, L. Himanen, A. Ladines, D. Sikter, M. Nakhaee, Á. Fekete, T. Chang, A. Golparvar, J. Márquez, S. Brockhauser, S. Brückner, L. Ghiringhelli, F. Dietrich, D. Lehmberg, T. Denell, A. Albino, H. Näsström, S. Shabih, F. Dobener, M. Kühbach, R. Mozumder, J. Rudzinski, N. Daelman, J. 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).

New members in FAIRmat team!



Ron Hildebrandt
Development team
Area B



Sarthak Kapoor
Development team
Area A



Carolin Rehermann
Communication and outreach expert
Area F



Yaru Wang
Materials informatics expert
Area E



Huayana Terraschke
Task leader A3
Area A



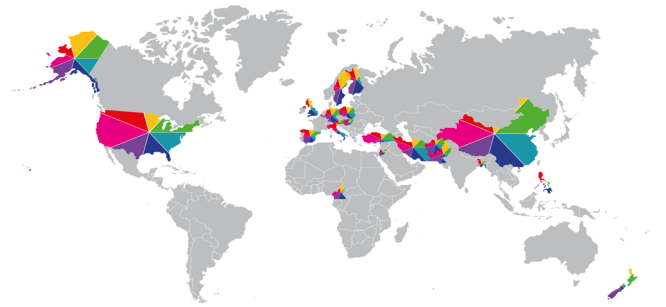
Tristan Bereau
Task leader C2
Area C



Kevin Jablonka
Task leader E7
Area E

Team Overview

FAIRmat offers a stimulating, multidisciplinary, and highly diverse working environment with ample development opportunities and flexible working hours. Apply now to join our team of experts in Berlin and across Germany!



Detailed World Map Vectors by Vecteezy

FAIRmat coworkers come from across the globe!

To see current job openings, visit our [website](#).

Stay in touch

Consortium FAIRmat,
c/o Humboldt-Universität zu Berlin
Zum Großen Windkanal 2, 12489 Berlin, Germany

❄ Website: <http://www.fairmat-nfdi.eu>

❄ Twitter: [@FAIRmat_NFDI](https://twitter.com/FAIRmat_NFDI)

❄ LinkedIn: [company/fairmat-nfdi](https://www.linkedin.com/company/fairmat-nfdi)

❄ Email: fairmat@physik.hu-berlin.de