



FAIRmat Newsletter

VOLUME 5 | July 2024

Editorial



Welcome to the fifth FAIRmat Newsletter!

FAIRmat is now about halfway through its first funding period. It is time to ask "Where do we stand?". Our project and users meetings in June amazed many of us. However, are we already that far? We have good concepts, ontologies, and functioning and productive software solutions. Everything is coming together. Everything is progressing, even if some gaps are still visible. However, it has also become clear that we may soon run into a problem of scale: we have to serve an ever-increasing circle of interested parties with the same core team. And the entire community is much larger still. We will certainly have to adapt our working methods and our decisions. We have to question ourselves: "What has priority?" The request of an interested project or the solution to a software problem?

What also became apparent, we have a community that is contributing more and more:

- interested scientists who adopt our concepts and make contributions, be it parsers or device drivers;
- device manufacturers or CAMELS experimentalists, both offering a FAIRmat-compliant output;
- collaborative research centers that further develop our concepts;
- students who handle electronic lab notebooks

The number of collaborators is growing, enabling us to progress even faster.

This acceleration is already noticeable! I am looking forward to the next 2.5 years and think to myself: this is going to be exciting!

Heiko B. Weber

Co-spokesperson and leader of Area B: Experiment

FAIRmat news

New communication channel: Discord

In line with the NFDI recommendation, we left the social media platform X, formerly Twitter, in April 2024. We are now focusing our social media communication on [LinkedIn](#) and have started a [Bluesky](#) account. For interactive and casual communication with our users, we have set up a [NOMAD Discord channel](#). Several sub-channels are active to focus the respective communication on specific topics. If you have not already done so, join our Discord channel and get into contact with our users and experts.

LLM hackathon

FAIRmat hosted a successful on-site satellite event of the [LLM \(Large Language Model\) Hackathon for Applications in Materials and Chemistry](#) at the Center for the Science of Materials Berlin (CSMB) on May 6 - 8. Around 50 participants from Friedrich-Schiller-Universität Jena (FSU), Bundesanstalt für Materialforschung und -prüfung (BAM), Helmholtz-Zentrum Berlin (HZB), Humboldt-Universität zu Berlin, and the Max-Planck-Gesellschaft joined us. The hackathon was kicked off by Kevin Jablonka, who gave an introductory talk followed by a lively discussion.

The event showcased innovative applications of LLMs in materials science. Participants engaged in a variety of projects and presented their work, and some of them took part in the global competition. Here, you can find a [collection of submitted projects](#).

We are thrilled to announce that our colleague Bernadette Mohr was a member of the prize-winning team "LangSIM". The project pitch is available [online](#). Our co-workers [Nathan Daelman](#)



and [Hampus Näsström](#) were part of teams awarded the Anthropic prize. We are impressed and delighted by how much can be achieved in just three days during a focused hackathon.

New faces join FAIRmat leadership

FAIRmat has recently implemented a new task within [Area C \(Theory and Computation\)](#), focusing on wave-function based methods. We welcome our new task leader, Denis Usvyat, to the FAIRmat team.

Being Area leaders for quite some time already, FAIRmat now welcome Heiko B. Weber and Silvana Botti also as co-spokespersons and, consequently, the Friedrich-Alexander-Universität Erlangen-Nürnberg and Ruhr Universität Bochum as co-applicant institutions.

Over 13 million data entries in NOMAD

FAIRmat has reached a significant milestone, with [NOMAD](#) surpassing 13 million published entries. These entries represent research on over 3 million different materials, with the total size of the uploaded files exceeding 100 terabytes.

By automating the detection of file types and extracting data and rich metadata, NOMAD makes all these entries findable and searchable. It also enhances their reusability, as all the entries align with a common interoperable data schema. These aspects allow aggregating data in a common format, even if the raw data sources are diverse. Additionally, this enables NOMAD to provide enhanced functionality built on top of the interoperable data model.

Project milestones

NOMAD measurement plugin

The NOMAD measurement plugin integrates with the NOMAD platform to handle experimental materials data. As a community-driven tool, it benefits from collaborative development and user contributions. Initially focused on X-ray diffraction (XRD), it provides a data schema for generating FAIR data for 1D and 2D XRD scans. A file reader supports formats from different vendors, namely Rigaku (.rasx), Malvern Panalytical (.xrdml), and Bruker (.brml). This plugin provides a common data structure, automatic metadata extraction, and visualization of XRD scans, eliminating the need for vendor-specific software. Based on NOMAD's Base Sections, it provides tools like a workflow visualizer and instrument history to enhance the user experience. In addition, the data can be analyzed with various tools available in the NOMAD ecosystem. More methods will be added into this plugin in the future to expand its capabilities. Users can install it on their NOMAD Oasis or test it on the [Example Oasis](#).

For more detailed documentation and usage instructions, visit the [NOMAD measurement plugin page](#).

NOMAD CAMELS - the newest member of our NOMAD family

Our configurable desktop software for controlling measurements and collecting FAIR experimental data had recently two reasons to celebrate: the [release of version 1](#) is out in the world, and [our paper on NOMAD CAMELS](#) is published in the Journal of Open Source Software (JOSS).

Try it out and benefit from our latest developments: [CAMELS](#) can now automatically upload measurement data to NOMAD, NOMAD Oasis, and eLabFTW. We added support for more than 200 scientific instruments through the open-source instrument driver model 'SweepMe!'. Finally, many other improvements enhance the user experience and functionality, such as improved plotting speed of live data, organizing measurement protocols in tabs for a clear overview, optional write protection for measurement protocols, and many more.

NOMAD catalysis app

One of FAIRmat's use cases is heterogeneous catalysis. We have developed a plugin with an [app](#), which helps to publish catalysis data in a FAIR way in NOMAD and simplifies data aggregation and retrieval. The app acts as a search-filter and visualization tool at the same time. All entries that populate the catalysis-specific results section through their normalizer show up in this app. It is possible to filter for specific reactions, reactants, and/or products. One can also start from the catalyst perspective and filter by elemental composition or preparation method. Reaction conditions can also be used as filter. The app's highlight are interactive scatter plots, which display actual catalyst properties such as conversion and selectivity. Customization allows the user to further select which data points are displayed, such as specifying reactant and product, and to choose which system property to be used as marker. Clicking on a data point leads the user directly to the original data entry.

FAIRmat guide - Legal Aspects in Research Data Management

The main purpose of the [guide](#) is to raise awareness of the legal aspects of research data management (RDM). It provides practical advice on legal considerations that are important for researchers in condensed-matter physics and materials science in Germany. The guide is tailored to be used by researchers from all experience levels.

The guide covers a range of legal aspects, including contract law, intellectual property rights such as database rights, copyright, and licenses for research data and code. It also addresses regulations on international data transfer, open access, data preservation mandates, cyberse-

curity measures, and research ethics. Understanding and adhering to these legal aspects facilitates data sharing and reuse, thereby the overall advancement of science.

By offering practical advice and references to original legal documents and laws, the guide helps researchers comply with legal requirements throughout their research activities. Although it is not a legal reference, the information in this guide has been carefully gathered and subsequently reviewed by RDM lawyers. Key takeaways include the importance of institutional and contractual policies for data preservation, respecting third-party intellectual property rights when sharing data, and adhering to open-access requirements.

The FAIRmat infrastructure



Markus Scheidgen
Infrastructure coordinator

NOMAD plugins and distributions

FAIRmat uses a bottom-up approach to build NOMAD, an extensible and customizable data infrastructure. To customize and extend NOMAD, users and data stewards can build NOMAD plugins. We offer several plugin entry points to provide support for new data types (schemas), files types (parsers), processing and analysis (normalizers), specific search interfaces (search apps), example data, and much more.

Writing plugins requires knowledge of NOMAD internals and some programming skills. To enable our community, we offer several options to learn. First, we have two example uploads on NOMAD that provide detailed notebooks with explanations that take users through all steps of writing a schema, writing a parser, processing data, and adding customized visualizations. These notebooks have been used in two FAIRmat tutorials already. Second, we updated our [plugin documentation](#) with details on all different plugin entry points. And third, we created example projects and templates, which make it very easy to quickly scaffold new plugins. While writing plugins is the first step, the second step is to actually add plugins to NOMAD.

We are now introducing NOMAD distributions that allow administrators to create customized NOMAD installations that run only the plugins they need. A NOMAD distribution is a special git project, where administrators edit a file that lists all desired plugins. From there, the build process is fully automated via GitHub or GitLab CI/CD pipelines. We plan to provide one central official NO-

MAD distribution that includes all the official NOMAD and FAIRmat plugins. This central distribution will also act as a directory for contributed third party plugins.

Since plugins and distributions exist as individual software projects, they can be developed independent of NOMAD itself, allowing a wider community to contribute to the NOMAD ecosystem. We expect that this will greatly help us to establish a long-standing community and support for a sustainable NOMAD.



Meet our users



Ta-Shun Chou
Junior researcher
Leibniz-Institut für Kristallzüchtung (IKZ)

What is the research focus in your group?

After a one-year period as a postdoctoral researcher at the IKZ (Leibniz-Institut für Kristallzüchtung), I am in a transition phase to become an independent researcher and to fund a startup together with IKZ. Currently, I am compiling research and business proposals, and depending on their outcome, I hope to start my next step soon. My research from the last years is focused on the epitaxial growth of Ga₂O₃ via MOVPE (Metalorganic Vapor Phase Epitaxy). One of my main achievements is to develop a MOVPE growth process of high-quality Ga₂O₃ films for high-power applications and reveal the growth mechanism. Overall, my research is based on the following three pillars:

- deploying MOVPE growth process for high-quality film
- utilizing in-situ reflectance spectroscopy for monitoring film growth
- implementing ML (machine learning) to systematically understand the growth parameters to control, e.g., film morphology, instead of a trial and error approach.

What challenges do scientists face when applying the FAIR principles to their research data?

One of the main challenges is the “conflict of data”. Scien-

tists still determine how much of their data to share and often want to protect them, leading to non-shared data. I usually experience that scientists tend to share only “old” data that has already been published in citable articles. I am often in a dilemma and would like to share my data according to the FAIR principles. However, the common interests of protecting the success of research conflicts with data sharing.

Specifically, achieving the ‘I’ in FAIR (Interoperability) is a significant hurdle in the field of MOVPE. Even when synthesis parameters are shared, reproducibility in another laboratory remains uncertain due to factors like the use of different MOVPE machines from different vendors. Overall, NOMAD helps me to share data easily and makes data meaningful to other scientists.

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

Based on my experience, research data management (RDM) varies from one group to another and is often personalized. The individual conducting most experiments typically determines the workflow of how and even if RDM is carried out. As a result, RDM practices within a research group tend to evolve over time. Depending on the scientist, they may vary, leading to potential data loss and knowledge loss. Essential data sets or parameters are often unknown or shared with other scientists.

With the help of FAIRmat and the providence of machine learning, I try to transfer all data to NOMAD. In the long run, NOMAD will give us a more structured way of managing data over long periods. Thus, NOMAD ensures the maintenance of knowledge by a standard data structure.

Using NOMAD makes data sharing easy and allows us to use machine learning to improve synthesis parameters. This inspired me to focus on and implement machine-learning approaches in my research.

Meet our experts



Julia Schumann
Heterogeneous catalysis expert in Area E: Use Cases

What is it that you do here at FAIRmat?

I am a domain specialist for heterogeneous catalysis, one of the use cases within FAIRmat. I develop schemas, work on visualization and search of catalysis data in NOMAD and coordinate with NFDI4Cat regarding standardization efforts in catalysis, such as the vocabulary Voc4Cat.

What drew you to join FAIRmat?

I joined FAIRmat because I like to work at the interface between different specialist domains such as experimental research, modeling, and IT. Here, I can combine my experience in experimental heterogeneous catalysis with my expertise in computational catalysis and additionally improve my coding and IT skills. I think it is high time to transform the precious data that accumulate during experiments into structured formats that are machine readable and can be used for machine learning. With the NFDI initiative and in FAIRmat, we have the opportunity to develop practical tools, workflows, and standards that can really make life easier for materials scientists, so that RDM is not a burden.

What is your favorite thing about working at FAIRmat?

I strongly support FAIRmat’s mission to make research data more useful and FAIR, and the people I interact with are very helpful and friendly.

Opinion article



Michael Krieger
Task leader in Area D: Infrastructure

FAIR experimental data collection

Experiments in solid-state physics and materials science often consist of complicated and individually composed setups. Every intricate detail matters and could influence the acquired data and, thus, the result of the experiment. These descriptive data are the metadata, which are all important to understand or even reproduce experiments. In reality, however, metadata are frequently documented only incompletely, sometimes just as part of the file name or directory structure with cryptic abbreviations, sometimes hand-written in paper lab notebooks, and sometimes metadata remain only in the memory of the scientist – not a good basis for sustainable data!

For example, the main data of interest in a spectrally resolved photoluminescence measurement of a material are intensity (or photon count) as a function of wavelength. This is what will be plotted and discussed in a graph. But there are many other details that need to be known to understand the data, such as: Which excitation was used? At least wavelength and power are required, but also polarization, focus, filters and lenses, etc. How was the spectrum recorded? Spectrometer? Grating? CCD camera or photomultiplier, or in general, which instruments were used with which settings? What was

the sample temperature during the measurement? And so on. Without these details, the experimental dataset is incomplete, cannot be understood, and is, therefore, not FAIR.

Luckily, scientists do not have to collect all those data by hand. In principle, computer software can easily do it. Modern experiments are controlled by computers anyway. However, individually composed experimental setups require measurement programs that have been developed specifically for that experiment. Such software, often written by the scientists themselves, frequently acquire (almost) only the raw data. This means that most of the metadata are lost and – in the best case – need to be tediously documented manually later on.

To overcome this shortcoming without any additional efforts, FAIRmat has developed the open-source software tool [NOMAD CAMELS](#). It is a configurable desktop software that controls your experiments. CAMELS automatically records all computer-accessible and relevant data and metadata during the experiment and stores them in a structured hdf5 file, which uses NeXus standards where applicable. These data files are fully self-descriptive and even contain the human-readable and the computer-executable measurement protocol. Anyone can understand or reproduce the data, because all the details are known. Curious about NOMAD CAMELS? Then try it out. It not only supports you at data acquisition, but also at measurement protocol definition and instrument communication. This is why we say that NOMAD CAMELS is your simple path to FAIR experimental data.

More information [here!](#)



What is: AAI?



Felix Dietrich

Coordinator in Area D: Infrastructure

Authentication and Authorization Infrastructure (AAI) is a crucial component of modern digital security, ensuring that users are who they claim to be (authentication) and that they have permission to access specific resources (authorization). AAI can involve access management with passwords, biometrics, multi-factor authentication, and access control lists to manage and secure access to data and systems. AAI is vital in preventing unauthorized access, data breaches, and ensuring compliance with regulatory standards. A robust AAI system not only protects sensitive information but also builds trust with users and stakeholders. For the central installation of NOMAD, we mostly use the Keycloak software for AAI. It is possible to change this to a local solution for NOMAD Oasis installations.

Collaborations with other NFDI consortia

Physical Sciences in NFDI

[PSinNFDI](#) is now a well-established collaboration among the [NFDI](#) consortia working in the natural sciences. For the second time, we have organized an event titled “FAIR Data Principles in Physical Sciences in NFDI - Workshop on use cases and FAIR data”. During this workshop, FAIRmat presented a shortened version of its introduction to the FAIR principles, using a LEGO® example for illustration. [DAPHNE4NFDI](#), [PUNCH4NFDI](#), and [NFDI-MatWerk](#) presented their tools to create FAIR data workflows. A recording of the workshop is available [online](#).

NFDI communicators meeting in Braunschweig

In addition to monthly online meetings, the communicators of all NFDI consortia met in person on May 7 - 8, 2024, in Braunschweig at the Helmholtz-Zentrum für Infektionsforschung. Possible directions for a joint NFDI communication strategy were discussed during two world cafés on the first day. On the second day, presentations and workshops on the topics of “Science communication” and “Wikiversum” took place. Carolin Reherrmann, communication expert in Area F: Training, represented FAIRmat in the meeting and jointly organized the session on “Science Communication”.

Collaboration with NFDI4Cat

To enable machines and computers to understand the data we store in our databases, we must first establish agreement among researchers on the specific meanings of terms, concepts, and quantities. This is done through definitions of vocabularies and ontologies, which must be developed within the research communities. For our use case heterogeneous catalysis, we have a natural point of contact with our sister consortium [NFDI4Cat](#), who have set up a [pipeline in Github](#) to facilitate the community-driven development of a vocabulary called [Voc4Cat](#). We are involved in discussions with the developers and have started contributing to Voc4Cat with the addition of new concepts. We use the international resource identifiers (IRIs) issued for the concepts to link the catalysis-related terms (i.e., quantities and sections) in NOMAD to achieve a semantic enrichment of our data. We have also contributed to a [white paper](#) by NFDI4Cat titled "Ontology-based Data Management and Interoperability for Catalysis".

Outreach and training

FAIRmat seminars series

Our successful FAIRmat seminar series was continued with two talks: Giovanni Vignale visited us on May 14, and gave a talk about "Geometric Density Functional Theory". The week after, on May 23, we welcomed Janosh Riebesell, who talked about "Foundational Machine Learning Potentials - Challenges and Opportunities". Both talks, taking place in Berlin, have also been broadcast online and will soon be available on our [FAIRmat / NOMAD YouTube channel](#).

DPG Spring Meeting of the Condensed Matter Section (SKM) 2024

Multiple co-workers from FAIRmat participated in the DPG Spring Meeting of the Condensed Matter Section at Technische Universität Berlin from March 17 - 22, 2024. Our colleagues from Area C, Theory and Computation, gave several talks on the principles of FAIR data and how to work with NOMAD in the field of simulation and computational materials science. During the "Hacky Hour", which was all about practical tools for daily scientific work, our latest member of the NOMAD family, NOMAD CAMELS, was introduced to the audience. After last year's success, we were present together with our sibling consortium, DAPHNE4NFDI, in the exhibition area with an information booth about research data management and NOMAD. The FAIRmat contributions were rounded off with a tutorial on "FAIR research data - generation, handling and analysis within the FAIRmat infrastructure".

Fourth FAIRmat Users Meeting

Our 4th FAIRmat Users Meeting was held on June 13 - 14, 2024, and for the first time outside of Berlin at the Friedrich-Alexander-Universität Erlangen-Nürnberg. We had a rich program with a panel discussion on the next steps in research data management, interactive sessions with user talks and hands-on workshops on RDM education, getting started with NOMAD, tailoring NOMAD to people's needs, and our new tool NOMAD CAMELS for FAIR data collection. In this meeting, our new comic characters, Lisa, Thomas, and Henry were introduced, who represent our users and collaborators. We plan to further develop storytelling around these personas to improve community building and the FAIRmat communication strategy.



Group picture of the participants in the Fourth FAIRmat users meeting at the Friedrich-Alexander-Universität Erlangen-Nürnberg.

Lange Nacht der Wissenschaften (LNdW) 2024

On June 22, 2024, we participated in the "Lange Nacht der Wissenschaften 2024" at the Center for the Sciences of Materials Berlin (CSMB) in Berlin-Adlershof with other research groups from the center. This is an annual event centered around science communication and public engagement. We contributed with 2 stands to bring across the fundamentals of programming and the importance of research data management (RDM) through a gamification approach based on LEGO®.

Upcoming events

FAIR-DI European Conference on Data Intelligence 2024

Building on the success of previous conferences, the FAIR-DI European Conference on Data Intelligence 2024 will take place in Karlsruhe, Germany, on October 27 - 30, 2024. [Here](#) are all the details on the conference topics, program, and invited speakers. Registration is open until August 15, 2024, via this [link](#).

Stay up to date

Our well-known FAIRmat tutorial and seminar series will continue in early autumn. To stay up to date with our upcoming events, check the FAIRmat website, join our mailing list, or follow us on our social media channels. You find all details below in the "Stay in touch" section.

Selected shots from the last few months

Welcome to our visual retrospective of the past few months! We have been involved in fantastic activities, and what better way to commemorate them than with a collection of photos? Dive in and relive the highlights with us!

We look forward to exciting future events and can not wait to create more unforgettable memories together!



*FAIRmat project meeting - June 11, 2024.
We are strengthening bonds and collaborations during the poster session.*



*LLM hackathon - May 6, 2024.
Pizza delivery to fuel our programming creativity!*



*Social evening - June 12, 2024.
Celebrating our success with great colleagues, cold drinks, and good vibes.*



*After-work board-game session - June 5, 2024.
Finishing a productive work day with board games.*



*Hands-on RDM education workshop - June 14, 2024.
Professors and scientists revisiting student lab courses.*

New members in FAIRmat team!



Fabian Schöppach
Development Team
Area E



Levi Mattes Bürki
Student Assistant
Area F



Bernadette Mohr
Multiscale Modeling Expert
Area C



Vikrant Chaudhary
Development Team
Area E



Lev Ginzburg
Development Team
Area B



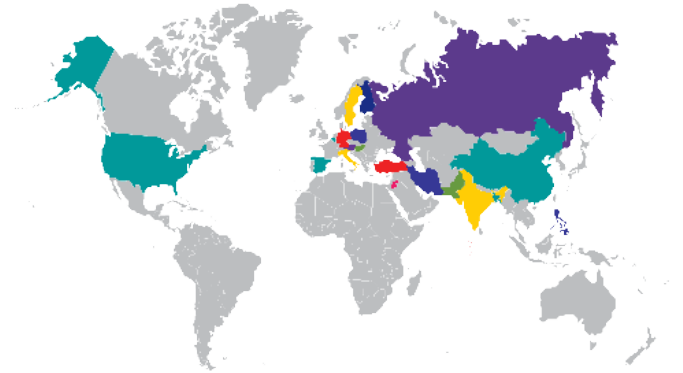
Denis Usvyat
Task Leader C4
Area C

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

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

Team overview!

FAIRmat coworkers come from across the globe!



Recent publications

- ❄️ A. D. Fuchs, J. A. F. Lehmeyer, H. Junkes, H. B. Weber, and, M. Krieger, *NOMAD CAMELS: Configurable Application for Measurements, Experiments and Laboratory Systems*, J. Open Source Softw. 9, 6371 (2024).
- ❄️ M. Baldovin, A. Browaeys, J.M. De Teresa, C. Draxl, F. Druon, F. Fradenigo, J.-J. Freffet, F. Lépine, J. Lüning, L. Reining, P. Salières, P. Seneor, L. Silva, T. Tschentscher, K. van Der Beek, A. Vollmer, and A. Vulpiani, *Matter and Waves, Chapter 3 in EPS Grand Challenges - Physics for Society in the Horizon 2050*, Ed. C. Hidalgo, IOP Publishing (2024).
- ❄️ M. Kuban, S. Rigamonti, and C. Draxl, *MADAS: A Python framework for assessing similarity in materials-science data*, preprint (2024).

Stay in touch

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

- ❄️ Website: fairmat-nfdi.eu
- ❄️ LinkedIn: linkedin.com/company/fairmat-nfdi/
- ❄️ Email: fairmat@physik.hu-berlin.de

NOMAD Discord!

Join our brand new NOMAD Discord, where you can connect with fellow researchers, share ideas, and get answers.

