FAIRmat Newsletter

VOLUME 7 I AUGUST 2025

Editorial

Welcome to the seventh FAIRmat Newsletter!

The past months have been pivotal for FAIRmat. At the Interim-Report Symposium in Bonn, our reviewers acknowledged the project's significant progress, characterizing it as "already well prepared and highly relevant". They highlighted the project's growing influence on both the German and the international research-infrastructure landscape. Building on that momentum, we have finalized and submitted the follow-up proposal.

From a technical perspective, FAIRmat's progress is equally evident. Highlights include a prominent presence at the DPG Spring Meeting in Regensburg, the Tutorial Tour that has disseminated NOMAD training to numerous laboratories across Germany, and the release of the NOMAD catalysis plugin. Concurrently, the integration of Base4NFDI services – namely, IAM4NFDI for unified access management and Jupyter4NFDI for cloud-based, reproducible workflows – positions the platform as a comprehensive, end-to-end solution for the global community.

In this issue you will find highlights of our developments, upcoming events, and the people who make them possible. We are grateful to all authors and contributors and look forward to what lies ahead.

Martin Albrecht
Co-spokesperson and leader of Area A: Synthesis

FAIRmat news

FAIRmat at DPG Spring Meeting in Regensburg

From March 16 to 21, our team participated in the DPG Spring Meeting of the Condensed Matter Section in Regensburg. Colleagues from various FAIRmat Areas presented recent developments in NOMAD, led a successful tutorial session on Using NOMAD's Workflow Utilities to Improve Data Management and Facilitate Discovery in Materials Science, co-organized two symposia and – together with DAPHNE4NFDI – hosted a joint information booth at the exhibition for scientific instruments and literature.

The list of all contributions is available on our website.

LinkedIn and Discord: strong growth across channels

This summer, we surpassed 1,000 followers on our LinkedIn page. It has become our main platform for sharing project milestones, announcing events and training opportunities, and highlighting community activities. It connects us with researchers, institutions, and collaborators across disciplines. Just as exciting, our NOMAD Discord server, launched in January 2024, has grown to over 500 members. It has quickly become a vibrant space for technical exchange, instant user support, and collaboration. The server hosts discussions on NOMAD plugins, schema development, and research data workflows, all organized into dedicated subchannels.

This growth reflects FAIRmat's commitment to building both professional visibility and hands-on community support. Whether you follow us for updates or work with us on FAIR data solutions, we are glad you are here!

New training module: getting started with NOMAD

To make NOMAD more accessible to new users, we have introduced a dedicated training module designed to lower the entry barrier. It guides learners through essential tasks using the graphical user interface, such as exploring data, uploading files, creating datasets, assigning DOIs, and using NOMAD as an electronic lab notebook (ELN).

The module has already been applied in several contexts. For example, it was used during the live online hands-on tutorial in January (recordings available), in internal onboarding sessions for FAIRmat members and PIs, and as a key component of the NOMAD Tutorial Tour in May.

Structured as a step-by-step tutorial with real-world examples, the training materials are easy to follow and ideal for self-paced learning. You can access the training module at this link.

Project milestones

Integration with Base4NFDI: FAIRmat adopts IAM and Jupyter services

FAIRmat has integrated two Base4NFDI services – IAM4NFDI and Jupyter4NFDI – marking a significant step towards improving user experience, enhancing interoperability, and aligning with the OneNFDI strategy.

With IAM4NFDI, NOMAD now supports secure and federated login using institutional credentials or trusted providers, such as ORCID, GitHub, or Google, all supported via the Helmholtz AAI. This enables access for users from over 400 German institutions and many international ones.

FAIRmat has also utilized Jupyter4NFDI, a centralized JupyterHub service, to support its training materials and tutorials. A first use case complements the recent Chemical Society Reviews article "From Text to Insight: Large Language Models for Chemical Data Extraction", offering interactive AI examples in chemistry and materials science directly executable via Jupyter4NFDI.

These integrations underscore FAIRmat's role in connecting domain-specific research with national infrastructure services and in lowering barriers to reproducible, FAIR-aligned data practices.

base4nfdi Basic Services for NFDI

Base4NFDI lays the groundwork for a federated NFDI by providing core services, guiding the adoption of FAIR principles, and connecting communities across disciplines.

NOMAD catalysis plugin official release

The NOMAD catalysis plugin is now fully integrated into the central NOMAD platform, offering new capabilities for managing and publishing catalysis data in a FAIR-compliant manner. This plugin introduces three key components designed to enhance the process of catalysis research data management:

- Schema package: a set of dedicated schemas to structure information on catalyst samples and catalytic measurements. These ensure consistent formatting, improve data quality, and support standardized research workflows.
- Catalysis app: an interactive dashboard to search, filter, and visualize over 1,500 catalysis data entries. Users can filter data by reaction, catalyst characteristics, or conditions, and visualize key properties (e.g., conversion vs. selectivity) through interactive scatter plots. The catalysis app can be accessed on the central NOMAD via this link.
- Example upload: a guided example illustrating how to prepare and upload catalysis data to NOMAD.

Together, these components streamline catalysis data workflows and promote standardization and reuse.

For more details on installation, usage, and functionality, you can explore the full documentation here.

The FAIRmat infrastructure

Lauri Himanen
Technical coordinator of Area D: Infrastructure

NOMAD release cycle

Earlier this year, we introduced a more predictable release schedule for the NOMAD software, which now includes release candidates (RCs) ahead of each stable version. We aim to publish a new release every 2-3 months to provide users with regular and reliable updates. To stay informed about upcoming releases and important updates, we encourage you to join our Discord server, where we share announcements, release notes, and engage with the community.

Before each stable release, we publish one or more RCs, which are tested in our beta deployment and evaluated by external Oasis administrators. This early feedback loop helps us identify and resolve potential issues before they reach production, while also giving the community a chance to contribute to the development and stability of the NOMAD platform.

Improved NOMAD deployment documentation and template

Cloud platforms provide a convenient and consistent environment for self-hosting NOMAD. The software is designed for easy deployment on cloud infrastructure, and we have expanded our documentation to include step-by-step instructions for setting up NOMAD on Amazon Web Services (AWS). We also plan to add examples for other cloud providers in the near future.

To help you quickly configure a custom NOMAD deployment, we provide a GitHub template that streamlines the setup process. The template now includes detailed instructions for enabling SSL encryption to secure communication with your Oasis instances. We have also added automated tests that verify the correct processing of example uploads and ensure that the Jupyter image used in Oasis starts up properly.

In addition, we are working on extending the template to support NOMAD deployments using Kubernetes. Kubernetes enables scalable installations by distributing workloads across multiple virtual machines, each of which can be configured with specific hardware resources. This is essential for supporting a growing user base and accommodating resource-intensive tasks or remote tools that require specialized hardware.

Upcoming improvements for NOMAD apps

NOMAD apps are used to build custom faceted search interfaces that target specific areas of research, materials or methodology.

We are adding improved validation for apps that will notify developers early on of any potential issues with the app configuration, such as missing or invalid quantities or other incorrect configuration values. In the future, quantity definitions for apps will be loaded dynamically from the server instead of being declared statically at the app creation time. This will make it possible to build applications "on-the-fly" and greatly simplify the app development cycle.

Meet our users

Edgar R. Nandayapa
Postdoctoral researcher,
Helmholtz-Zentrum Berlin

In the fast-paced world of energy research, collecting data is not the hard part, it is making them useful. With the rise of machine learning and mandates from funders like the European Union to meet the FAIR principles, effective research data management has become a top priority.

At Helmholtz-Zentrum Berlin, we are tackling this challenge head-on with the NOMAD infrastructure. We aim to help scientists document their work in a way that benefits their daily research activities and the broader scientific community by providing them with useful tools.

From spreadsheets to smart workflows

When we introduced NOMAD, one of our first hurdles was getting it adopted. Many scientists found it difficult to enter data directly into the web interface. At first, we supported them with a dedicated data steward, but this approach was not scalable because they relied on continuous assistance. So, we changed course.

Since most researchers already use spreadsheets to plan experiments, we closely collaborated with them to develop a spreadsheet template that feeds directly into NOMAD. It is compatible with our metadata parsers and dramatically lowered the entry barrier. Suddenly, scientists could upload their data independently, and they did.

Turning metadata into motivation

Even with easier tools, keeping track of metadata is still a time-consuming task. That is why we focused on providing value to researchers. We created two types of Python-based applications:

- Tools that analyze data stored in the NOMAD server, such as spectral or JV measurements, offering immediate insights and lab-wide comparisons.
- Standalone apps that assist with calculations and experiment planning.

These tools are open and editable by any team member. This fosters a collaborative environment where researchers can build on each other's work, making NOMAD more than just a data repository. It is a platform for innovation.

What is next?

Today, NOMAD is an integral part of our workflow. We are constantly evolving guided by user feedback and real lab needs. Our mission is clear: to make FAIR data not just a requirement, but a valuable asset in every lab. Want to get involved or learn more? Reach out! We are always looking for new ideas and explore potential collaborators.

Meet our experts

Esma Birsen Boydas
Domain expert in Area C: Computation

What motivated you to join FAIRmat?

Throughout my bachelor's, master's, and PhD studies, I tackled a wide range of molecular modeling projects at six different research institutes around Europe, sometimes even developing new methods from scratch. Yet each project lived in its own niche: different codes, file formats, and analysis tools. I saw firsthand how much time and effort went into translating results from one group's toolkit to another's. When I learned about FAIRmat's mission – to make data and tools truly interoperable across disciplines – it felt like the perfect next step. My background in method development seemed tailor-made for ontology work: I understood both the scientific details and the practical pain points of producing, sharing, and analyzing complex data. Joining FAIRmat gave me the chance to help build a common language for the community, so that future researchers can spend less time wrestling with formats and more time making discoveries.

What is your role at FAIRmat?

As a domain expert in Task C4: Wave-function Based Methods, I work on integrating quantum chemistry and quantum technologies into NOMAD. My background is in theoretical chemistry – I completed my PhD at Humboldt University's Chemistry Department before moving over to FAIRmat. My daily focus is on designing and refining data structures that capture both quantum chemical simulations and the metadata needed for quantum information workflows. I spend most of my days extending the nomad-simulations schema, so it can accommodate standardized inputs and outputs for everything ranging from molecular energy computations to quantum computing experiments. A fun part of my routine is seeing how a tiny tweak in the schema opens up new possibilities for automating data analysis or visualizing results in novel ways.

What is your favorite thing about working at FAIRmat?

Two things stand out. First, the interdisciplinary spirit: I get to collaborate with experimentalists synthesizing novel compounds, theorists dealing with quantum algorithms, and computer scientists, building the infrastructure to handle it all. That mix keeps every week and every meeting fresh and exciting. Second, I love having a "bird's-eye view" of science – helping to build a platform that aims to serve not just quantum chemistry, but essentially every subfield in the natural sciences. It is incredibly rewarding to know that the ontologies and tools we develop today could streamline research across many diverse fields tomorrow.

Opinion article

Laurenz Rettig
Task leader in Area B: Experiment

Modular metadata: a smarter way to capture experimental context

Metadata are the unsung heroes of experimental science. Without them, raw data is just numbers – uninterpretable and often unusable. Metadata provide the crucial context: what was done, how, with what settings, and under which conditions. And yet, despite their importance, metadata collection is often tedious, error-prone, and inconsistently practiced.

Fortunately, many metadata can be acquired automatically. Tools like NOMAD CAMELS make it possible to capture all machine-accessible parameters during the experiment – from instrument settings to sample environment – and store them directly in a structured HDF5 file following NeXus conventions[1]. This means less manual logging and fewer gaps in context.

But what about the metadata that cannot be pulled from an instrument? The so-called "soft" metadata – like sample composition, preparation method, or reasoning behind measurement sequences – still require human input. Additionally, certain parameters simply may not be available in an automated fashion in every lab due to older, manual hardware. Interoperability and reusability require capturing such metadata in a structured, semantically organized Electronic Lab Notebook (ELN). However, as experiments become more and more complex, such entries multiply, and filling an ELN quickly becomes a bottleneck: researchers either repeat the same entries endlessly, or resort to copy-paste, with the inevitable risk of inconsistency.

One solution to this challenge is to rethink how we structure metadata entries in ELNs. Instead of logging everything in one massive form, we can break metadata into a hierarchy of modular, linked entities organized by how frequently they change. This way, a scientist only has to update those entries that actually vary between measurements, while other values link automatically to their most recent values. Take a temperature-dependent study measured across different excitation wavelengths as an example. Depending on the setup, the experiment might

vary temperature per wavelength or vice versa. Instead of re-entering all settings for each measurement, the user could link a new entry to a given wavelength setup and update only the temperature field. All other metadata – instrument configuration, sample ID, or environmental controls – remain linked to their most recent versions. It is an approach that mirrors how scientists think during experiments and adapts to real-world workflows.

During data ingestion and analysis, these linked entries can be traversed to assemble a complete metadata record, automatically resolving references to produce, e.g., full NeXus-compatible datasets. This not only reduces user burden but also ensures that metadata completeness does not come at the cost of usability.

Structured ELNs like elabFTW[2] or NOMAD's ELN[3] already support this kind of hierarchical, linked structure. When used thoughtfully, they help bridge the gap between automation and human input – allowing scientists to focus on experiments without sacrificing the integrity of their data.

Good metadata should not feel like a chore. By aligning structure with scientific practice, we can make FAIR data not just achievable, but intuitive.

References

1. A. Fuchs *et al.*, J. Open Source Softw. **9**, 6371 (2024)
2. N. CARPi *et al.*, J. Open Source Softw. **2**, 146 (2017)
3. M. Scheidgen *et al.*, J. Open Source Softw. **8**, 5388 (2023)

Outreach and training

Over the past six months, FAIRmat has actively engaged with the community through a variety of events, introducing new formats and strengthening key collaborations.

Our training activities took a new turn with the Tutorial Tour (May 19-22), where we taught NOMAD across multiple locations. At the same time, the FAIRmat tutorials series continued with Episode 16, and the FAIRmat seminar series hosted talk by Prof. Helge Stein and Dr. Janine George.

Collaboration within the NFDI consortia remains strong. As part of the Physical Sciences in NFDI (PSinNFDI), we co-organized two joint colloquia, featuring Prof. Sören Auer and Prof. Isao Tanaka – both events drew wide audiences. Recordings are available on our YouTube channel.

We continued presenting NOMAD on the international stage. At the European Materials Research Society (E-MRS) Spring Meeting (May 26-30) in Strasbourg, we not only contributed scientific talks, but also showcased FAIRmat at the exhibition and promoted NOMAD at the Industrial Forum.

In recent weeks, we further connected with the community at two major events: the Sixth FAIRmat Users Meeting and the Long Night of the Sciences – Berlin. The Users Meeting took place at Ruhr-Universität Bochum on July 2-3 featuring two days of talks, tutorials, and networking focused on AI- and ML-driven materials science, as well as the application of FAIR data principles for "AI-ready" research data.

Our participation in the Long Night of the Sciences – Berlin provided a unique opportunity to engage with the younger generation – and perhaps inspire them to pursue a path in academia.

Your continued interest and engagement drive us forward. We welcome your ideas, feedback, and participation in our events. To stay informed, visit the FAIRmat news page for detailed updates on our activities.

Upcoming events

We will be present at several upcoming events across Europe, offering talks, workshops, and opportunities to connect in person:

- 2nd DPG Fall Meeting, September 8-12, 2025 – Göttingen, Germany
- On-site LLM Hackathon in Materials and Chemistry, September 11-12, 2025 – Berlin, Germany
- International Conference on Silicon Carbide and Related Materials (ICSCRM), September 14-19, 2025 – Busan, Korea
- Open Databases Integration for Materials Design (OPTIMADE) Workshop, September 15-19, 2025 – Vilnius, Lithuania
- Machine-actionable Data Interoperability for the Chemical Sciences (MADICES 3) Workshop, October 20-24, 2025 – Villigen, Switzerland

To stay informed and never miss an event, check our events page or follow us on LinkedIn. We are continually adding new opportunities related to NOMAD education and community building.

New members in the FAIRmat team!

Dea Bello
Assistant
Area G

Haoyu Yang
Development Team
Area D

Juan David Salamanca
Development Team
Area D

Recent publications

- A. Mirza, N. Alampara, S. Kunchapu, M. Ríos-García, B. Emoekabu, A. Krishnan, T. Gupta, M. Schilling-Wilhelmi, M. Okereke, A. Aneesh, M. Asgari, J. Eberhardt, A. M. Elahi, H. M. Elbeheiry, M. V. Gil, C. Glaubitz, M. Greiner, C. T. Holick, T. Hoffmann, A. Ibrahim, L. C. Klepsch, Y. Köster, A. Kreth, J. Meyer, S. Miret, J. M. Peschel, M. Ringleb, N. C. Roesner, J. Schreiber, U. S. Schubert, L. M. Stafast, D. Wonanke, M. Pieler, P. Schwaller, K. M. Jablonka, A framework for evaluating the chemical knowledge and reasoning abilities of large language models against the expertise of chemists, Nat. Chem. **17**, 1027 (2025).

- T. Denell, L. Himanen, M. Scheidgen, and C. Draxl, Automated identification of bulk structures, two-dimensional materials, and interfaces using symmetry-based clustering, npj Comput Mater **11**, 25 (2025).

The full list of our publications and invited talks can be found on our website.

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
- Bluesky: @fairmat.bsky.social

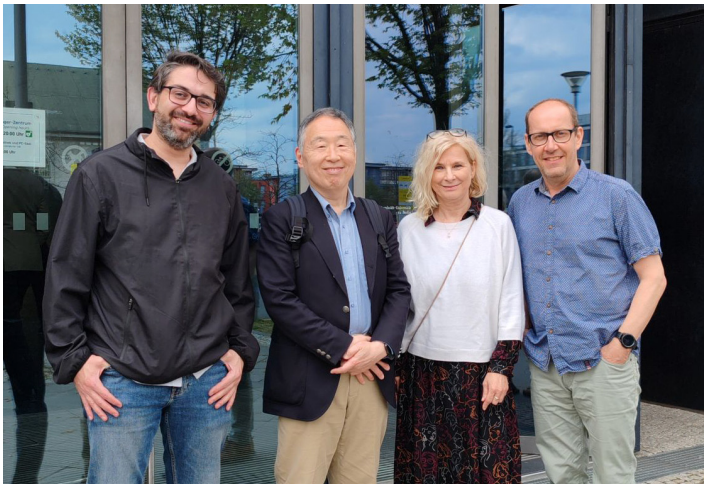
NOMAD Discord!

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

Selected shots from the last few months

Our team at the FAIRmat Project Team Workshop in Eisenach.

Collaborative energy in action.



Prof. Isao Tanaka with FAIRmat members at PSinNFDI Joint Colloquium.

Engaging with the community at E-MRS Spring Meeting booth.

Drawing a full house at DPG SKM Spring Meeting tutorial.

The FAIRmat outreach team at DPG SKM Spring Meeting.