



# FAIRmat Newsletter

VOLUME 3 | JUNE 2023

## Editorial



Welcome to the third FAIRmat Newsletter!

You will again find a number of reports on exciting progress in different areas of our consortium. As the Leader of Area E "Use Case Demonstrators" I would particularly

like to highlight the recent workshop "Accelerated Discovery of New Materials", which took place as a WE-Heraeus-Stiftung funded conference in Bad Honnef from May 15-18.

Even the Deutsche Bahn realized the importance of this conference - the train strike was canceled at the last minute so that all attendees could arrive in time for a very exciting meeting. The conference was opened with a plenary talk by our Chairperson, Claudia Draxl, and then proceeded with a number of invited and contributed presentations. In fact, the interest in the conference was so great that a number of interested colleagues could not be admitted. Quite a large number of attendees from the FAIRmat consortium joined this meeting.

From conversations with some attendees who were initially rather skeptical about the importance of research data management in experimental science, I learned that in particular the adoption of Large Language Models (LLM) like ChatGPT motivated them to reconsider. For more details, see the report on page 7.

*Christof Wöll, co-spokesperson and leader of Area E*

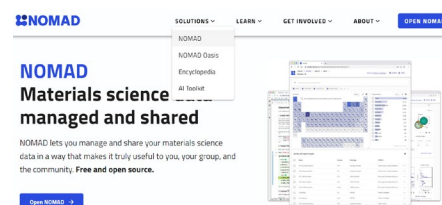
## FAIRmat news

### FAIRmat at the DPG SKM spring meeting

The FAIRmat team participated in a variety of ways at the [DPG Spring Meeting of the Condensed Matter Section](#) in Dresden on March 26-31. FAIRmat's presence was marked by a special plenary talk discussion on "NFDI and FAIR research data," featuring esteemed panelists. Additionally, we organized a focus session on making experimental data FAIR and set up an exhibition booth in collaboration with fellow NFDI consortium DAPHNE4NFDI. Through these activities, FAIRmat succeeded in raising awareness about FAIR data management, sharing experiences, and engaging with the community. The community's active engagement with FAIRmat demonstrated the growing importance of the FAIR principles in materials research and emphasized the significance of open and accessible data for scientific advancements.

### A new website and logo for NOMAD

The new NOMAD website, launched in March 2023, features a sleek and modern design



that is easy to navigate and provides a comprehensive overview of NOMAD's unique features and benefits. The updated design incorporates sections that showcase the solutions offered by NOMAD, the latest infrastructure updates, FAIRmat tutorials and the various channels available for user support. Visit the NOMAD website at <https://nomad-lab.eu/>.



## Photoemission spectroscopy community and technology partners agree on a standard application definition

Photoelectron spectroscopy researchers from several research institutes and technology partners from industry came together for FAIRmat's second "community meets technology partners" workshop on May 11-12. We had a broad discussion on the general process of standardization and reviewed in detail the application definition for multidimensional photoelectron spectroscopy, NXmpes, which is being developed within FAIRmat. The application definition has already been accepted by the NeXus International Advisory Committee (NIAC) and released as a contributed definition. Further suggestions for improvements are still being collected and the discussions will continue.

## Project milestones

### Integration of electronic lab notebook software and formats

Following the bottom-up approach of FAIRmat, we integrate with existing electronic lab notebook (ELN) software, similar to other data types. We define schemas to capture the data model for various ELN software, and provide parsers or processing tools that import data from those ELNs. We have already implemented this for LabFolder and eLabFTW and are currently looking at Chemotion. FAIRmat is now establishing actual use-cases for LabFolder and eLabFTW.

### A practical guide to writing a data management plan

FAIRmat has released an informative brochure titled "FAIRmat guide to writing a research data management plan" to aid scientists in condensed-matter physics and materials science. The guide assists researchers in complying with the DFG's requirements and ensuring proper research data handling. It is available [online](#) under the CC BY 4.0 license and has also been distributed in printed form at conferences. The brochure was praised by the community and the DFG. It has been downloaded over

250 times and approximately 150 printed copies have been circulated. This resource shows FAIRmat's commitment to promoting proper research data management practices and supporting researchers in effectively managing their data.

### NOMAD's interactive search dashboards

The interactive search dashboards (or NOMAD Apps) allow one to define a domain- or application-specific view of a dataset. Each application contains a customizable dashboard composed of widgets for tailored data interaction. The first published application of this kind is the [Solar-Cell App](#), integrating more than 40,000 entries from the Perovskite Database Project. Its customizable dashboard, consisting of interactive widgets, allows for searches, navigation, and visualization of solar-cell data using numerous properties and performance metrics as filters.

## The FAIRmat infrastructure

### NOMAD Plugins

*Markus Scheidgen, infrastructure coordinator*



Extensible and customizable data management software is an important success factor for FAIRmat. If we want to organize many different types of data from heterogeneous research domains and for all kinds of workflows, NOMAD has to be adapted for different environments before it can be used to its full potential. For the computational materials science domain alone, NOMAD supports over 60 different file formats, and each format requires additions to NOMAD's schema, its own parser, different normalization steps, and potentially specialized visualizations. Building support for all kinds of file formats and workflows into a monolithic software package, as we are doing now, is no longer sustainable. Therefore, we will separate core NOMAD functions from individually (re-)usable NOMAD plugins.



NOMAD already provides different means to extend and customize. For example, users can upload schemas that describe their new data structures, they can customize the search, or add tools in the NOmad Remote Tools Hub (NORTH) including Jupyter notebooks for data analysis. However, there is currently no way to add functionality to NOMAD's processing, like new parsers or normalization routines. Consequently, it is hard for Oasis users to add support for new data types. On the other hand, if support for all kinds of formats is built into NOMAD by default, Oasis installations are crowded with functionalities which are not needed by the people using them. Furthermore, a monolithic NOMAD prevents external developers to contribute support for further data types.

During the 9<sup>th</sup> [FAIRmat tutorial](#) (April 26), we presented NOMAD's new plugin mechanism. This mechanism allows plugin developers to create schemas and parsers without running a NOMAD data infrastructure, lowering the entry barrier for adding new functionality. Developers can fork one of our template GitHub projects and create their first plugin in a matter of minutes. Plugins can be developed and tested without a running NOMAD and with fast development roundtrips. Plugins can be mounted into Oasis installations, meaning that Oasis administrators can extend the official NOMAD releases instead of building their own custom NOMAD versions.

Currently, the first FAIRmat partners are evaluating plugins and we have started to convert existing schemas and parsers to plugins to test the mechanism within FAIRmat first. This will improve how we work within FAIRmat, as domain experts and Areas can work independently within the clear boundaries defined by plugins. The challenge for the near future will be to build a vibrant ecosystem of plugins contributed from not just FAIRmat, but FAIRmat collaborators and beyond. To this end, we plan to provide a plugin registry and process that will allow contributors to publish and share their plugins. More sophisticated options to distribute and install plugins will

be developed as well. Further, we want to extend plugins from parsers to normalizers and custom APIs.

## Meet our users



*Prof. Dr. Carmen Herrmann  
Institute of Inorganic and Applied Chemistry,  
Universität Hamburg*

### What is the research focus in your group?

We are mainly working on the first-principles simulation of systems for molecular electronics and spintronics, with methods ranging from density-functional theory to molecular dynamics simulations, to postprocessing analysis (e.g., for evaluating electrical conductance, exchange spin coupling, or Kondo properties) and to machine learning models. We are applying and evaluating existing methods as well as developing new ones. Two of the central questions we are interested in are understanding the mechanisms underlying chiral induced spin selectivity and magnetoresistance in molecular junctions.

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

Deciding on how to assign maximally useful metadata and, initially, deciding on appropriate sharing platforms and licensing models (that is, the "activation barrier" is still high).

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

Our goal is to have all raw data easily accessible within the group (and for collaborators), as soon as projects leave the initial "explorative" stage, and to share the raw data (and analysis scripts, if applicable) underlying publications with the community once the paper is out. The FAIRmat consortium has kindly provided us with a NOMAD Oasis, which has already helped new students to get started, in particular by enabling them to retrieve optimized structures and data on calculations by those who had already left the group.



## Meet our experts



*Dinga Wonanke, domain expert in Task E5*

### What is it that you do here at FAIRmat?

I work as a developer in Area E (use case demonstrators for Metal Organic Frameworks, MOFs). The overarching goal of Area E is to create a fully interactive database that will expedite the rational design and synthesis of high-performing stable MOFs. My primary role is to build a comprehensive MOF database that maps the crystal structures of all experimentally synthesized MOFs to their experimental synthetic conditions and properties. I also implement new tools to search, manipulate, visualize, and extract these data. Finally, I assist various end-users by providing them with curated data and explanations to their unprecedented experimental outcomes.

### What drew you to join FAIRmat?

My goal as a researcher is to design new tools for predicting the optimal conditions of chemical reactions, which can successfully lead to a paradigm shift from the conventional cost-ineffective and environmentally unfriendly trial-and-error synthetic approach. This desire was first nested in me during my Ph.D. wherein I implemented a new metric for predicting a type of light-driven reaction. However, at the end of my studies, I realized that although reactivity metrics are powerful enough to predict the outcome of reactions, they do not provide synthetic chemists with the right level of detail for optimal synthesis. I then moved to the U.K. to do a postdoc where I realized that the main bottleneck in predicting optimal synthetic conditions is the unavailability of a comprehensive and structured database of experimental synthetic conditions.

I joined FAIRmat because it offered me the golden opportunity to begin fulfilling my dreams, through the construction of a fully comprehensive database that maps structures to their properties and experimental synthetic conditions.

### What is your favorite thing about working at FAIRmat?

I sincerely enjoy my job at FAIRmat because the vision and goals of FAIRmat fully align with my main research goals. My job is like my hobby, and I am genuinely passionate about it. Moreover, FAIRmat has one of the healthiest working environments. The team spirit is exceptional because everyone seems to be one call away from providing any help I need. In addition, I get to team up with a bunch of friendly people whom I genuinely enjoy spending time with.

On a personal level, working here has made a significant improvement in my programming skills through coding and getting helpful feedback from peers in the team. I have also improved my communication skills, through presenting at various conferences and outreach programs.

At a professional level, working here has impacted me positively through the creation of new research ideas and the establishment of a significant number of collaborators.



## Opinion article

### Implementing ELNs in physics practical courses

*Heiko B. Weber, leader of Area B and Task D5*

*Michael Krieger, leader of Task D5*

How can we make the whole of physics a discipline that stores data in an AI-ready manner, thus advancing the discipline along the fourth paradigm? The plethora of tasks on the way to a FAIR data structure cannot be solved by the NFDI consortia alone. The residual burden lies with individual scientists, especially the youngest generation, who need to be trained to adapt existing concepts to their particular case. However, when talking to colleagues about changing the canon of physics to add data literacy, one quickly realizes that the scope of the course is capped in ECTS credits, so something that was previously considered essential would have to be eliminated. It will not be easy to organize a consensus here.

At the university of Erlangen, we have recently introduced electronic lab notebooks (ELNs) into the existing third-semester lab courses (and there are similar initiatives at TU and HU Berlin). At this stage, our students already have base skills in Python programming from a mandatory course in the first semester. The introduction of ELNs does not incur any costs in ECTS calculation, does not require any changes in the examination regulations, and yet proved to be a game changer. The students accepted ELNs immediately and appreciated working in teams of two with a shared ELN instead of having two notebooks. They quickly learned to deal with logs, data, and (ideally scripted) evaluations within the same platform, similar to our planned NOMAD Oasis concept. This has implications for the whole department: once students begin doing independent research for their bachelor thesis, they introduce or promote ELNs to their respective workgroups. In this way, they become driving forces for the culture change that consortia like FAIRmat aim to achieve. With their Python and data skills, they are better prepared than

any student generation before to actively contribute to data-driven science.

In addition to all this, we as the organizers of the lab course have an added advantage: We now have a sandbox at hand to try out new ELN concepts. How can we best configure an ELN? Should we specify things further or should we leave it open? Semester after semester, we can experiment with our ELN concepts to gain experience. After three lab courses with openBIS, our next experience will be with NOMAD Oasis ELNs.



Heiko B. Weber



Michael Krieger

### What is: an ELN?

*Sebastian Brückner, Area A coordinator*

An Electronic Lab Notebook (ELN) is a digital platform designed to replace traditional paper-based lab notebooks, streamlining the documentation, management, and sharing of experimental data within the materials-science domain. ELNs facilitate interdisciplinary collaboration, enabling scientists to create, store, and access data in a structured, machine- and human-readable format. This enables efficient data retrieval, interpretation, and reproducibility of research findings.



ELNs offer various features such as templates, version control, and metadata tagging, which allow for better organization and contextualization of experimental information. Additionally, they integrate with other scientific software and analytical tools, promoting seamless data exchange and analysis.

In materials science, ELNs enhance the discovery and optimization of new materials by capturing critical data on synthesis and characterization. Their interoperability with



databases and machine learning algorithms accelerates the materials design process and expedites scientific breakthroughs. Overall, ELNs contribute to the digital transformation of materials science, promoting data-driven research and innovation.

## Collaborations with other NFDI consortia

### Physical Sciences in NFDI

We have joined forces with fellow [NFDI](#) consortia [DAPHNE4NFDI](#), [MaRDI](#), [NFDI-MatWerk](#), [NFDI4Cat](#), [NFDI4Chem](#), and [PUNCH4NFDI](#) to form the group Physical Sciences for NFDI. We unite experts on a broad spectrum of topics in physics, chemistry, mathematics and informatics. Our goal is to foster collaboration among the physical sciences consortia and host joint events such as workshops or summer schools. We currently host a series of talks in which we invite leading scientists to showcase good data practices to an international, interdisciplinary audience. Our most recent speaker was [Taylor Sparks](#), associate professor at the University of Utah, Royal Society Visiting Wolfson Fellow at the University of Liverpool and host of the [Materialism podcast](#).

Physical Sciences  
in [nfdi](#) 

## Outreach and training

### Second FAIRmat users meeting



The second FAIRmat users meeting took place on June 7, in Berlin. This highly anticipated public event attracted a

diverse audience including users, collaborators, interested researchers, and representatives of other NFDI consortia. The meeting provided an invaluable platform for knowledge sharing, networking, and fostering collaboration within the community. The event commenced with an engaging introduction to FAIRmat presented by FAIRmat's co-spokesperson, Christoph T. Koch. This was followed by an inspiring invited talk delivered by [Prof. Carmen Herrmann](#), a user of NOMAD Oasis. The FAIRmat experts then presented a series of talks on research data management and NOMAD use cases. All the talks were made accessible to a wider audience through live Zoom broadcast and can now be viewed [on our YouTube channel](#). The event's program also included a poster session held in the foyer of the IRIS building. This dynamic session featured 25 informative posters showcasing ongoing developments in FAIRmat, other NFDI consortia and recent results from RDM practitioners. Participants had the opportunity to engage in lively discussions, exchange ideas, and forge new connections while exploring the diverse range of topics presented.

### FAIRmat and NOMAD documentation workshop

This May, members of the NOMAD development team and FAIRmat domain experts met with documentation experts from [Canonical](#) for an interactive workshop on the [Diátaxis](#) framework for writing documentation. The core concept of Diátaxis is to split documentation into distinct categories based on whether it should share practical or theoretical skills, and whether it should aid study or active work. Our team very quickly got to work identifying how we can make the NOMAD documentation more effective and easier to use and will implement these ideas over the coming months!





## 787. WE-Heraeus-Seminar in Bad Honnef

The 787. [WE-Heraeus-Seminar](#) "Accelerated Discovery of New Materials", which took place from May 15-18 at the Physikzentrum in Bad Honnef, brought together around 70 researchers of all career stages and various backgrounds in physics and chemistry from 40 institutes in 12 countries. The program included 12 invited talks, 9 contributed talks, and 37 posters. These excellent contributions covered a wide variety of topics highlighting the importance of automation, high-throughput screening, modeling and digitalization in various areas of materials science, including electrocatalysis, renewable energy, and electronic device applications. These topics represent the cutting-edge research areas and emerging trends in the field. A somewhat unexpected – for most of the attendees – aspect of the meeting was the impact, which large language models (LLMs) like ChatGPT have already gained in materials science. This seminar provided a valuable and novel opportunity for researchers from various fields to exchange on the development and application of digitalization in materials discovery.



## Booth at the E-MRS spring meeting

After the success of our first exhibitor's booth at the DPG spring meeting, FAIRmat also had a booth at the [European Materials Research Society \(E-MRS\)](#) spring meeting in Strasbourg. This was a fantastic opportunity to meet the international materials-science community as well as other exhibitors to talk about FAIR data and NOMAD.

## Upcoming events

### CoRDI

With the first edition of the Conference on Research Data Infrastructure from September 12-14, 2023, the association German National Research Data Infrastructure (NFDI e.V.) is initiating a conference that will focus on establishing interdisciplinary research data management (RDM). Under the theme "Connecting Communities", national and international stakeholders from all research fields as well as from the infrastructure sector are invited to present their contributions to an excellent RDM of the future and to exchange information about the latest developments. [Find out more.](#)

### CECAM Workshop

On September 25-27, FAIRmat and the [Molecular Simulation Design Framework](#) (MoSDeF) group are holding a CECAM flagship workshop to highlight efforts towards FAIR data management for molecular simulations and to discuss standardization of metadata and interoperability within the community. The speakers represent a range of perspectives including FAIR-data-type projects and consortia as well as developers of simulation engines. There will also be proposals for metadata schemas/ontologies as well as research talks focusing on curating and usage of large datasets. [Find out more.](#)

## Recent publications

- ✳ M. Krieger, H. B. Weber, and C. van Eldik, *Früh zur Datenkompetenz*, Physik Journal 21 Nr. 12, p.42, 2022
- ✳ C. P. Marshall, J. Schumann, and A. Trunschke *Achieving Digital Catalysis: Strategies for Data Acquisition, Storage and Use*, Angew. Chem. Int. Ed. e202302971, 2023.



## The FAIRmat team

### Welcome to our new coworkers and PIs!



Uday Gajera  
Domain expert  
Area E



Ahmed Ilyas  
Domain expert  
Area D



Mehrdad Jalali  
Domain expert  
Area E



Sascha Klawohn  
Development team  
Area D



Siamak Nakhaie  
Documentation/training expert  
Area F



Lukas Pielsticker  
Domain expert  
Area B



Julia Schumann  
Domain expert  
Area E



Julian Siegmann  
Student assistant  
Area F



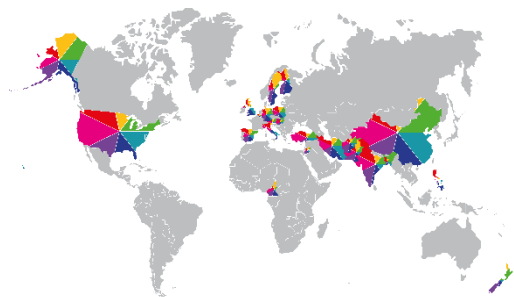
Erdmann Spiecker  
Task leader B1  
Area B



Adrianna Wojas  
UI/UX expert  
Area D

### Join the team

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!

We are currently looking for

- ✳ Data expert (biophysics / life science)
- ✳ Materials synthesis data scientist
- ✳ Optoelectronic materials data scientist

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

### Stay in touch

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

Twitter: [@FAIRmat\\_NFDI](#)

Email: [fairmat@physik.hu-berlin.de](mailto:fairmat@physik.hu-berlin.de)

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