Claudia Draxl, Heiko B. Weber

FAIRmat – democratizing data-enhanced research

Imagine, you have got excellent optical spectra after changing your crystal growth protocol. Now you want to compare the results with data from five years ago that were measured by a student who left your group. In most cases, this will impose difficulties in finding the respective dataset, getting all the relevant background information that might (or might not) have been laid down in a handwritten lab book. Then you might experience that the two datasets cannot be compared one-by-one because they lack relevant details. Often it appears easier to redo the old experiment rather than trusting old data. What a pity!

Imagine, you want to be a pioneer in your field and introduce a data-driven approach by training a Large Language Model (LLM) with all the relevant results worldwide. You have to organize a mechanism that crawls through heterogeneous data, in various formats and standards distributed at many places. In physical chemistry, this endeavor appears impossible today.

In both cases, it would be desirable to have the data in a standardized representation that respects the FAIR principles (findable, accessible, interoperable, and reusable) [1]. More practically, a platform that efficiently organizes FAIR data would be helpful. It would be a place where you can find and access the data, where interoperability is invisibly organized, and where everything is prepared for data reuse – for data collections from your own lab as well as worldwide open-science data.

This is the mission of FAIRmat [2]: supporting condensed-matter physics and the chemical physics of solids with a single integrated, extensible, distributed FAIR Research Data Management (RDM) service. A core element of FAIRmat is its comprehensive data management platform NOMAD [3]. As an infrastructure with structured data and interoperability as primary design principles, NOMAD fosters the definition of rich data schemas, acquisition from a wide range of sources, transformation of raw files into structured and reusable data, data sharing across institutions, and scalable analysis. With NOMAD, you can collect, organize, share, analyze, and publish FAIR data across the scientific core areas synthesis, experiment, and computation with a rich scientific spectrum of domain-specific applications.

NOMAD is an open-source software [4] that comes in two flavors: NOMAD and NOMAD Oasis (see Figure 1). We start our description with NOMAD Oasis that can be installed in your local and protected environment – on your desktop as a standalone service, or at your institution as a cloud service. There, you can organize your group's data. It can serve as a working environment where every result, be it from experiment or theory, is immediately processed and stored. The software includes also lab notebook functionality. You can perform complex analyses of your data using NORTH, the NOMAD Remote Tool Hub. It enables users to run containerized scientific tools and Jupyter notebooks directly on NOMAD-hosted data via a web browser. You can give selective access to your collaboration partners and even run joint, federated NOMAD Oases networks.

Currently, we count more than 85 registered NOMAD Oases worldwide – and are aware that there are many more in use. You may customize your NOMAD Oasis according to your special needs, using NOMAD's plugin mechanism. There are highly specialized plugins for perovskite solar cells, for catalysis, and for battery research, as well as others for specifying local workflows and electronic lab notebooks. The plugin concept was introduced less than two years ago, and since then, the community has shared already about 100 plugins that can be used and further developed by anyone.

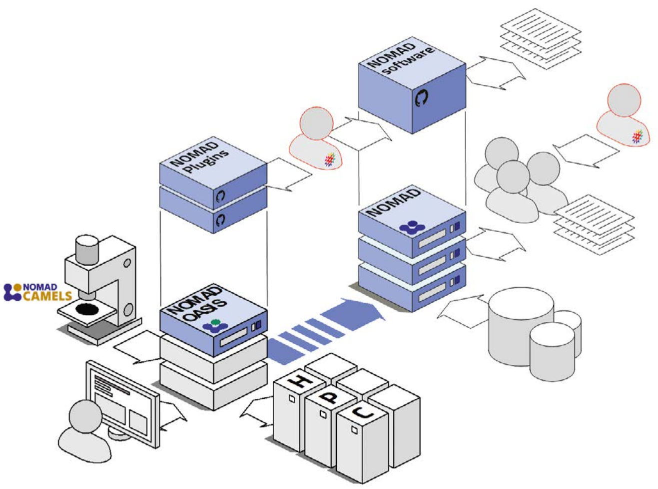

Fig. 1: The NOMAD ecosystem.

Prof. Dr. Claudia Draxl
Theoretische Festkörperphysik, Humboldt-Universität zu Berlin
Zum Großen Windkanal 2, 12489 Berlin
claudia.draxl@physik.hu-berlin.de
https://sol.physik.hu-berlin.de/

Prof. Dr. Heiko B. Weber
Lehrstuhl für Angewandte Physik,
Friedrich-Alexander-Universität Erlangen-Nürnberg
Staudtstr. 7 / Bau A3, 91058 Erlangen
heiko.weber@fau.de
https://www.lap.physik.nat.fau.de/

DOI-Nr.: 10.26125/5x3m-gy43

If you want to publish a dataset from your NOMAD Oasis, you simply transfer it to NOMAD. This is FAIRmat's central open-science platform, and in essence, the identical software. Here, you can by now find about 20 million complex entries, representing 4 million materials. For historical reasons, the vast majority of data in NOMAD are so far computational data. Upload of experimental data has started only recently. Similar to the NOMAD Oases, one can truly work on NOMAD with all its data. Here, you may compare world-wide research results and train AI models or LLMs. NOMAD, counting 4,000 visits per day from all over the world (see Figure 2), is the leading platform for analyzing materials data.

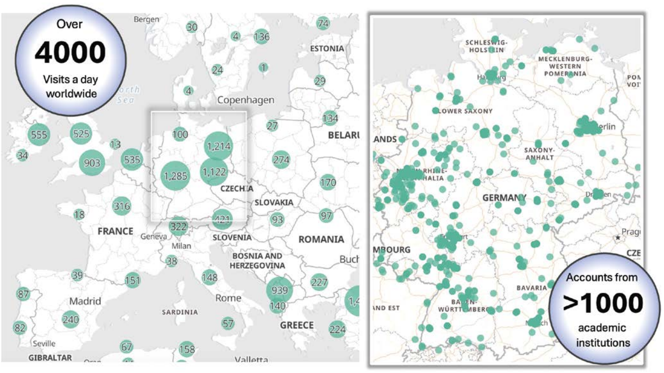

Fig. 2: NOMAD users from Germany and Europe-wide in July 2025.

Both NOMAD and NOMAD Oasis provide a rich portfolio of functionalities. For instance, the software is designed for meaningful domain-specific queries. To give an example (see Figure 3): You may specify a search for solar cells that have C_{60} as an electron conduction layer, further contain the element tin, and have an open-circuit voltage between 0.8 V and 1 V; and you may further filter the results by looking for data from a specific lab or from a certain timespan. If you have different needs, you may specify your own search app via the plugin mechanism.

NOMAD emphasizes interoperability, which is the most difficult aspect of FAIRness. On the computational side, for example, you may compare electronic structure from many different sources. Here, the results are predominantly obtained from density-functional theory (DFT), more recently also from wave-function-based methods. Meaningful comparison of physical quantities is possible because of the rigorous underlying data schema. When you upload data with a known data structure into NOMAD, a rich and highly-specific metadata set is immediately assigned. This applies, for example, to DFT data, where a large number of community codes is already supported through parsers. It also applies to experimental data that follow the international NeXus standard. In these cases, you don't have to do anything, the (meta)data are automatically stored according to NOMAD's underlying schema, Metainfo [5]. If, however, the format is yet unknown to NOMAD, upload is possible by providing the data schema or workflows together with the data.

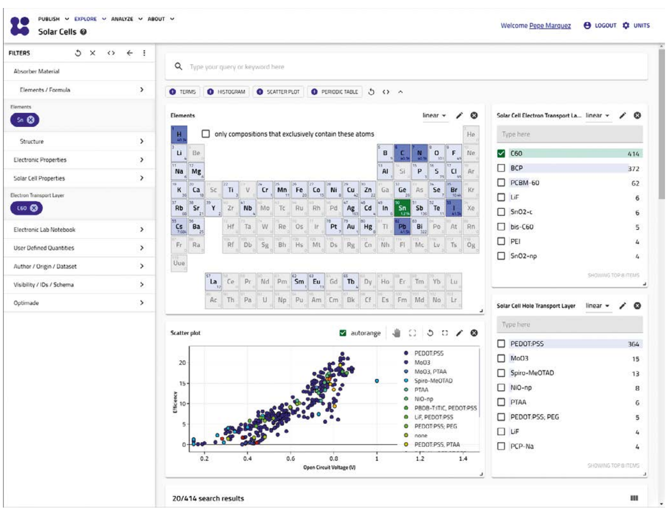

Fig. 3: Screenshot of a NOMAD search app. It is used for exploring experimental data of solar cells with various device parameters.

FAIRmat has developed "helpers", software that supports the researchers. One is pynxtools, which translates data from angle-resolved photoemission spectroscopy (ARPES), X-ray photoemission spectroscopy (XPS), various optical spectroscopies, atom-probe tomography, electron microscopy, and scanning-tunneling microscopy into the international NeXus format. With NeXus, data not only follow international community standards, they are automatically processed by NOMAD.

A valuable tool for orchestrating experimental labs is NOMAD CAMELS [6]. It is a software that has been developed to simplify the control of homebuilt experiments. Instead of spending weeks and months for writing a control software, an experiment can be organized with CAMELS within minutes or a few hours. CAMELS handles the communication with the instruments, and the scientist only needs to configure elements like a sweep, a loop, a PID control, or similar. CAMELS provides a graphical representation and organizes data evaluation. Most importantly, it writes the data in a standardized and highly structured way, including all available metadata, and cooperates with electronic lab books. Therefore, the data are immediately understood by NOMAD. With its application programmable interfaces (APIs), CAMELS is further developed as a backbone for lab automation.

The entire NOMAD ecosystem is open source. We experience valuable contributions from the community – within and outside Germany. To learn more about our developments and services, we recommend to follow our rich video documentation (more than 360 videos on our YouTube channel [7] with more than 1,500 followers). In particular, the young generation of scientists prefers to contact us via our Discord channel [8] to get into immediate contact with our software and domain experts. FAIRmat further attempts to bring data skills into the university curricula. Equipped with the necessary competencies, our students will become a driving force for dynamic, data-enhanced science.

This brings us back to our overall objective, which is to prepare for the fourth paradigm of science: Processing and collecting data shouldn't be an end in itself. A thoughtful and consequent handling of data gives immediate benefits to every researcher. Even more important is the emerging gain when structured data and AI come together. FAIRmat is committed to boost the cultural and technological change in solid-state physics and related disciplines.

We are happy about the opportunity to develop FAIRmat's ideas and solutions as part of the National Research Data Infrastructure (NFDI). Due to the NFDI and its 26 consortia covering a broad range of scientific disciplines, Germany is now in a leading position with respect to FAIR research data. FAIRmat consists of ~90 scientists (Principal Investigators, PIs) and about 35 employees. Our international PIs build bridges to similar efforts within Europe. We expect to be part of the European Open Science Cloud (EOSC) [9] soon.

The discussion how NFDI will be shaped beyond 2028 has begun. We appreciate the clear political commitment in favor of the NFDI becoming a permanent national institution. However, care has to be taken that this effort remains close to the requirements of the scientific domains rather than providing research data management as a discipline-agnostic, mere technical service. To keep up with the pace of the latest scientific developments, the infrastructure needs continuous adaption to ever-new methods, materials and functions, as much as the exchange with leading scientists of the respective disciplines. These community needs should be considered when designing the govermental structure of the future NFDI.

References

[1] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data*, **3** (2016) 160018.

[2] M. Scheffler, M. Aeschlimann, M. Albrecht, T. Bereau, H.-J. Bungartz, C. Felser, M. Greiner, A. Groß, C.T. Koch, K. Kremer, W.E. Nagel, M. Scheidgen, C. Wöll, C. Draxl, FAIR data enabling new horizons for materials research, *Nature*, **604** (2022) 635–642.

[3] NOMAD: http://nomad-lab.eu.

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

[5] 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. Himmer, S.V. Levchenko, M. Oliveira, A. Michalchuk, R.E. 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, M. Scheffler, Shared metadata for data-centric materials science, *Scientific Data*, **10** (2023) 626.

[6] J. Lehmeyer, A. Fuchs, M. Krieger, H.B. Weber, NOMAD - CAMELS - Configurable Application for Measurement - and Laboratory Systems, GitHub, 2021.

[7] NOMAD/FAIRmat YouTube Channel: https://www.youtube.com/c/TheNOMADLaboratory.

[8] NOMAD Discord: https://discord.com/invite/Gyzx3ukUw8.

[9] EOSC: https://open-science-cloud.ec.europa.eu/.

Prof. Dr. Claudia Draxl

Claudia Draxl is Einstein professor at the Humboldt-Universität zu Berlin. Her research interests cover theoretical concepts and methodology to gain insight into a variety of materials and their properties, with a focus on theoretical spectroscopy. In her group, the all-electron full-potential package **`exciting`** and the cluster-expansion package **`CELL`** are developed. She is spokesperson of the NFDI consortium FAIRmat. FAIRmat develops NOMAD, a research data management service for collecting, organizing, sharing, analyzing, and publishing FAIR materials science data.

Prof. Dr. Heiko B. Weber

Heiko B. Weber holds the Chair for Applied Physics at the Friedrich-Alexander-Universität Erlangen-Nürnberg. He performs experimental research addressing a broad range of fundamental phenomena in condensed-matter physics. In his investigations, he uses predominantly molecular materials, epitaxial graphene, and silicon carbide. He is co-spokesperson of the NFDI consortium FAIRmat, leading its Area "Experiment". His group is further leading the NOMAD CAMELS development.