FAIRmat Newsletter

VOLUME 2 I DECEMBER 2022

Editorial

Welcome to this latest edition of our newsletter! It will provide you with an overview of the project's progress and an outlook to its next steps. In this volume, we introduce several new recurring segments:

- In Meet our experts we introduce a member of the team who explains their role in FAIRmat and what motivated them to join us.
- In Meet our users we talk to a user about how they manage data in their research and their thoughts about the future of materials research.
- In What is? one of our experts explains a key concept in research data management in simple words.
- In the opinion section, one of our PIs shares their views on a key topic related to research data management.

Enjoy reading, explore our webpage and activities on social media, and don't hesitate to contact us if you would like to get actively engaged in FAIRmat.

Claudia Draxl, FAIRmat spokesperson

FAIRmat news

FAIRmat is one year old!

In October 2022, FAIRmat turned one year old! The first anniversary was celebrated by the launch of our new logo and informational website: www.fairmat-nfdi.eu/. It features a fresh design and comprehensive contents about FAIRmat's activities and infrastructure.

FAIRmat team gathers in Berlin for the project meeting

The second FAIRmat project meeting took place on November 14-16, 2022, at the FAIRmat headquarters in the Integrated Research Institute for the Sciences (IRIS Adlershof) of the Humboldt-Universität zu Berlin. Team members including principal investigators, FAIRmat domain experts, and external collaborators came together to report on achievements, deliverables status, progress and challenges, and to plan for the activities in the upcoming period of the project. The meeting offered an opportunity for discussions among each area, as well as inter-area discussions on topics of common interest and collaboration.

Community agrees on NXellipsometry

In September 2022 FAIRmat held the first workshop in the "Community meets technology partners" series, which aims to lay a foundation for users and vendors to work together to make experimental data FAIR. At this first meeting, members of the scientific community and the technology partners met in Leipzig and online to work together towards the goal of FAIR data handling in ellipsometry. Their very fruitful discussion was focused on reviewing a specific application definition, NXellipsometry, in detail. Read more online

Project milestones

FAIRmat experimental data models accepted as contributed definitions in the NeXus standard

A framework has been created to ease the creation of data models for different types of experiments. The models can be developed in a simple human-readable data serialization language (YAML files) and then converted automatically to the NeXus Definition Language (NXDL) which is presented to the community for feedback. More information

Over a dozen global NOMAD Oasis installations

The NOMAD software can now also be installed as a local data management system. Several research groups in materials science have already taken this opportunity, and the FAIRmat network is growing quickly.

Development of a general data model for four different synthesis methods

FAIRmat Area A developed a general data model for synthesis data which is based on the idea of a flexible data schema. Our data model serves the individual needs of different synthesis methods, in particular synthesis from the melt, gas phase, solid phase/solutions, and by assembly. We implemented our data model in NOMAD for different use cases where we made use of the newly developed NOMAD Oasis ELN feature.

The FAIRmat infrastructure

NOMAD: the software behind the FAIRmat data infrastructure

By Felix Dietrich and Markus Scheidgen, Area D


Felix Dietrich


Markus Scheidgen

To make data FAIR, we need to enable scientists to manage, share, and publish their data, explore available data, and use data in novel contexts (e.g. in artificial intelligence). Due to the nature of materials science and the plethora of available methods and tools, data are heterogeneous, inter-connected (e.g. by workflows), and not centrally governed. Also, at present there is no established culture of data sharing, curation, and publication.

To fulfill its mission, FAIRmat software needs incentives to properly manage data and curate metadata in the first place, and to follow a bottom-up approach that embraces data plurality and adapts to needs of local groups while also retaining enough generality to fulfill the needs of the materials-science community as a whole. FAIRmat develops the NOMAD software and maintains the NOMAD infrastructure based on this software.

NOMAD began in 2014 as a central repository for publishing input and output files of electronic-structure simulations. It implemented FAIR principles through the formalization and homogenization of data (common schema, structured data) through the NOMAD Laboratory Center of Excellence between 2016 and 2018. It underwent a major revision prior to FAIRmat, leading to a coherent software product. This included a first version of the NOMAD Oasis, i.e., the possibility to create independent NOMAD installations. However, NOMAD software still had a

narrow focus on publishing mainly DFT calculations, neglecting other aspects of data management and materials-science communities. At this point, NOMAD software was maintained by a small team of four developers.

During the first year of FAIRmat, the digital infrastructure team (Area D) extended the scope of NOMAD. The NOMAD Oasis had to become meaningful to non-DFT researchers and usable for managing local day-to-day data workflows beyond simulations. Therefore, we put a strong focus on the individual Oases including new features for data acquisition (parsers, NeXus, ELNs, Central Application for MEasurement and Lab-Supervision "CAMELS"), data organization (projects, folders, custom schemas), data analysis (by integrating notebooks and tools), and data governance (privacy, local sharing). During the first year we built a team of over 10 core developers and addressed over 500 issues in 1800 commits on the main software repository (not including parsers and other sub modules).

As laid out in the FAIRmat proposal, we follow an iterative approach with a major revision of NOMAD each year. We put the new features into the hands of scientists and collected feedback and experiences necessary to determine the direction for the coming year. The focus will now shift towards data sharing beyond a central installation, towards a federated network, centrally collecting metadata.

Meet our users

Prof. Lorenz Romaner
Chair of Physical Metallurgy and Metallic Materials, Department of Material Science, Montanuniversität Leoben, Austria

What is the research focus in your group?

The research activities in my group target the simulation of materials properties, in particular the design of crystallographic defects including dislocations or grain boundaries. A central aspect is to understand how changes in chemical composition affect atomic-scale quantities and their consequences for overall material behavior. A particular research focus is grain-boundary segregation.

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

There are several challenges. A big problem is related to missing metadata or context information, which makes it hard to extract meaningful information from data sets. In this sense, metadata standards would be highly beneficial, but they need first to be developed by the materials-science community. Another point is related to confidentiality issues in projects with industrial collaboration.

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

The general strategy is to use and develop workflows that enable us to standardize data acquisition and provide the necessary context information to make the data as reusable as possible. With the implementation of a NOMAD Oasis, we have the possibility to build a consistent database for grain-boundary segregation and to compare data and results from different collaborating groups.

Meet our experts

Nathan Daelman
Development Team, Area C

What is it that you do here at FAIRmat?

I am a domain expert in Area C (Theory and Computations). More specifically, I work on developing tools for the FAIR storage of data from ground-state calculations, which cover most of the current data in the NOMAD Archive. Together with the community, I am building a framework for addressing data quality (a much-requested feature) through the lens of basis sets and general numerical settings that are adopted for the stored

calculations. I am also working on an *ab initio* ontology in Area C, and aim to provide users with a knowledge map of all exchange-correlation density-functional approximations.

What drew you to join FAIRmat?

I come from a PhD and one-year postdoc right next to the Costa Brava in Spain, so it wasn't the weather that made me change. Jokes aside, I guess you could call it "frustration". As scientists, we ourselves and our environment tend to be single-minded in our focus on getting new, hot results. Once published though, the actual essence, the data, and its preservation become second thought. This mindset always comes back to haunt us. I've felt that, my fellow colleagues felt it, everyone does.

FAIRmat is among the leading projects in tackling data infrastructure. Because it has recognition and is open-source, I feel like it has a real shot at making materials-scientists lives easier and the overall output better. That's something I want to contribute to.

What is your favorite thing about working at FAIRmat?

I love the team. I can tell how enthusiastic everyone is about making data FAIR. While the wider community may still be skeptical about the feasibility of FAIR data, we have a whole group of people who decided to invest their career in this vision for the future of science.

The team is very diverse with respect to background and expertise, which leads to the most fascinating debates and discussions. Seeing how representatives from co-existing communities have such different approaches and viewpoints really puts my own understanding into perspective. Also outside of work, we have a lively bunch who are always up for a drink. Having only recently moved to Berlin, I'm very grateful for them.

Opinion article

Education of Research Data Management at Universities: We are failing our students!

By Erich Runge, Area F

We, a loose, informal group of physics professors with perhaps above-average interest in physics curricula, started thinking about implementing research data management (RDM) in university education about two years ago; and we started by listening to the community. As a first step, the German Physical Society DPG organized a workshop on "RDM in Physics Curricula." It quickly became clear that we are, frankly speaking, failing our students: Not only are we not teaching them RDM, but in many places there is even a lack of a concept for teaching programming, information and communication technology (ICT), and numerical skills. The fact that these three areas are usually seen as parts of the same thing only makes the situation worse. Similarly, simple data fittings in the basic lab courses do not prepare students for research data, which are worthless without recording metadata: Statistical analysis and data management are also not simply two sides of a coin.

Everyone knows that most physics graduates have to manage data and often also write computer code in their careers – but no one draws the necessary consequences. "You kind of learn that on the side" is what they often say. "But that's not how it should be", I say. Therefore, FAIRmat supports a project of the DPG to look at the topic "data in physics studies" in the broadest sense. A survey of the current situation should ultimately lead to possible educational scenarios and recommendations for action.

I am optimistic that soon at least lip service to FAIR data will be a natural part of all research proposals and soft-skill programs at our graduate schools. But I am doubtful whether our graduates will be able to form a meaningful sentence around the terms "ontology" and "metadata". Personally, I think this should become part of general higher education - if only to understand why Google works so well.

Whether you like it or not, the design of degree programs today takes place in an Excel-spreadsheet-dominated atmosphere of quality management and competence orientation. One good side of this is that my world view – formulated here in a very simplified and self-ironic way – "physicists can actually do everything" is being questioned. I don't want to discuss here who else has this world view - one or the other reader may be there. Well, I for one had way too much fun with my computational physics to realize that we actually should outsource the topics 'Introduction to ICT' and 'Programming' from the physics departments. Our colleagues in computer science can do some things better. Those who can program well will also be able to use our well-established legacy, e.g., F90 codes, the reverse is unfortunately not true.

Competence orientation means in a positive sense, preparation for later life. Thus, we must finally differentiate and clearly define the roles of programming, ICT and numerical skills as well as data understanding and data management in physics studies and map them in the curriculum. We owe this to our students.

What is: metadata?

By Sandor Brockhauser, Area B

Metadata is the contextualization and description of a data object, which allows for that data object to be properly interpreted and turned into information. FAIR principles require metadata to be rich enough to explain the meaning of the data at such a level that the represented information can be connected to already existing domain knowledge. Making it richer than necessary (e.g., explaining not only what an experimental data object is, but also why a particular experiment has been done) could even allow quasi-automatic generation of the technical parts, e.g., the methods sections, of our publications. Read more online

Collaborations with other NFDI consortia

FAIRmat continuously collaborates with the other consortia of the German Research-Data Infrastructure (NFDI) on cross-cutting topics. Part of this effort is the co-organization of the NFDI Physical Sciences Joint Colloquium, in cooperation with DAPHNE4NFDI, MaRDI, NFDIMatWerk, NFDI4Chem, NFDI4Cat, and PUNCH4NFDI. Together we bring the world's leading experts in research data infrastructure for physical sciences to our audience in Germany and around the globe.

In 2022 we had four speakers: James A. Warren (NIST), Chris Wolverton (Northwestern University), Nancy Washton (PNNL), and John R. Helliwell (University of Manchester). The series continues in the coming year, starting with Susanna A. Sansone (Oxford e-Research Centre) giving the talk "FAIR data: no longer optional, but it takes a village!" planned on January 9, 2023.

Outreach and training

FAIRmat users' meeting

The first FAIRmat users' meeting took place on November 16, 2022, as part of the FAIRmat project meeting in Berlin. The event offered an opportunity for current users, potential users, and scientists interested in research data management to get to know the activities of FAIRmat and meet and discuss with our experts. The event was well attended both on-site and in our live virtual room on Zoom.

The program started with an overview talk by the FAIRmat spokesperson Claudia Draxl, followed by presentations from each of the FAIRmat areas to explain the tools

and applications of FAIRmat. Invited speakers included our current users Lorenz Romaner from Montanuniversität Leoben and Michael Götte from the Helmholtz-Zentrum Berlin as well as prospective user Carlos-Andres Palma from the Chinese Academy of Sciences and Humboldt-Universität zu Berlin. All of the talks are now available to watch on our YouTube channel.

Left to right: Claudia Draxl, Lorenz Romaner, Carlos A. Palma, Michael Götte.

FAIRmat tutorials

During the first year of FAIRmat, we have held six hands-on tutorials. These highly attended online events offered current and potential users and members of the community a chance to learn about the tools and capabilities of the FAIRmat infrastructure, and how it enables a comprehensive research data management portal for FAIR data from condensed-matter physics and the chemical physics of solids. Real hands-on examples are demonstrated in an interactive manner. A list of past and upcoming tutorials can be found on our website, and videos are available on our YouTube and KouShare channels.

Upcoming events

FAIRmat at the DPG spring meeting 2023

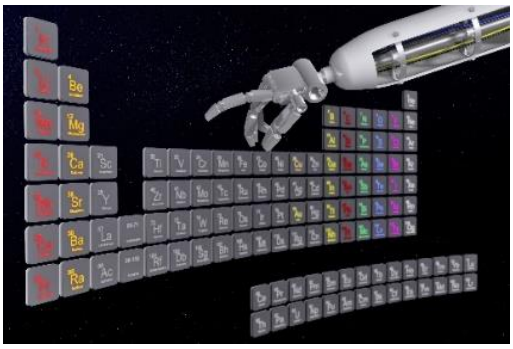
26 – 31 March, 2023

- We are organizing a focus session in the Surface Science division "Making experimental data F.A.I.R. – New concepts for research data management".
- We are hosting the "Diskussionsveranstaltung in der Mittagsschiene".
- Visit us at the FAIRmat booth for live demonstrations and advice on our tools!

FAIRmat WE-Heraeus-Seminar in Bad Honnef

15 – 18 May, 2023

- This WE-Heraeus-Seminar will bring together researchers at all career stages from experiment, theory, and computer science to discuss digitalization strategies for materials research.
- The seminar will demonstrate FAIR research with two novel classes of materials, e.g., high entropy alloys (HEAs) and metal organic frameworks (MOFs).
- Applications are open until March 12, 2023 on the WE-Heraeus Foundation website.



Recent publications

- M. Scheffler, M. Aeschlimann, M. Albrecht, T. Bereau, H.-J. Bungartz, C. Felser, M. Greiner, A. Groß, C. Koch, K. Kremer, W. E. Nagel, M. Scheidgen, C. Wöll, and C. Draxl, *FAIR data enabling new horizons for materials research*, Nature **604**, 635 (2022).
- L. Sbailò, Á. Fekete, L. M. Ghiringhelli, and M. Scheffler, *The NOMAD Artificial-Intelligence Toolkit: turning materials-science data into knowledge and understanding*, npj Comput Mater **8**, 250 (2022).
- M. Kuban, S. Rigamonti, M. Scheidgen, and C. Draxl, *Density-of-states similarity descriptor for unsupervised learning from materials data*, Sci. Data **9**, 646 (2022).
- C. Draxl, M. Kuban, S. Rigamonti, and M. Scheidgen, *Challenges and perspectives for interoperability and reuse of heterogenous data collections*, Section 4.1 in H. J. Kulik, et al. Roadmap on Machine Learning in Electronic Structure, Electron. Struct. **4**, 023004 (2022).
- M. Kuban, Š. Gabaj, W. Aggoune, C. Vona, S. Rigamonti, and C. Draxl, *Similarity of materials and data-quality assessment by fingerprinting*, MRS Bulletin **47**, (2022).
- Y. Luo, S. Bag, O. Zaremba, A. Cierpka, J. Andreo, S. Wuttke, P. Friederich, and M. Tsotsalas, *MOF Synthesis Prediction Enabled by Automatic Data Mining and Machine Learning*, Angew. Chem. Int. Ed. **61**, e202200242 (2022).
- M. Jalali, M. Tsotsalas, and C. Wöll, MOFSocialNet: *Exploiting Metal-Organic Framework Relationships via Social Network Analysis*, Nanomaterials **12**, 704 (2022).

Welcome to the FAIRmat family

The FAIRmat family has grown quickly in 2022, with two new PIs and 20 new staff members joining our team.

Miguel A.L. Marques
Task Leader (C1), Area C

Walid Hetaba
Task Leader (B3), Area B

Theodore Chang
Area D

Nathan Daelman
Area C

Thea Denell
Area D

Florian Dobener
Areas A and B

Carola Emminger
Area B

Adam Fekete
Area D

Amir Golparvar
Area D

Dinga Wonanke
Area E

Shuaihua Lu
Area C

Joseph Rudzinski
Area C

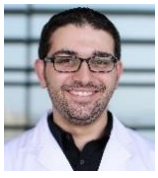
José A. Marquez
Areas A and E

Rubel Mozumder
Area B

Jonathan Noky
Area A

Hampus Näsström
Area A

José M. Pizarro
Area C



Ahmed Mansour
Area F

Alexander Welle
Area E

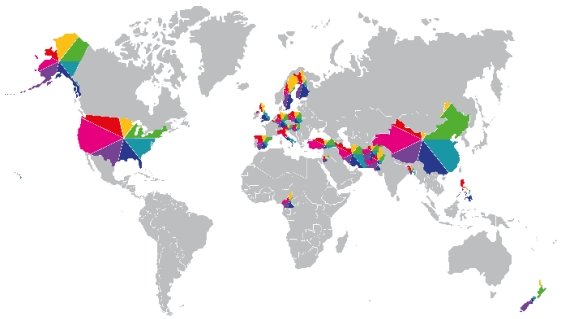
Lucia Rotheray
Area F

Naz Günes Pizarro
Area G

Sirkka Remes
Area F

Join our 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!

Current openings

- Data steward for Area E
- Software engineers

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