FAIRmat NEWSLETTER

EDITORIAL

A warm welcome to our brand-new FAIRmat newsletter!

The past 12 months have been very interesting and productive for FAIRmat, and 2020 is very busy and exciting already. In this first issue, we like to share with you some information about our recent, ongoing, and upcoming activities and the current status of the National Research Data Infrastructure (NFDI).

The NFDI Directorate is jointly established by the Karlsruhe Institute of Technology (KIT) and the Leibniz Institute for Information Infrastructure (FIZ). Its responsibility is to coordinate all consortia in tackling the challenges to work across borders and disciplines. FAIRmat very much welcomes the new NFDI Director, **York Sure-Vetter**. We are looking forward to a fruitful collaboration!

FAIRmat'S MISSION

Scientific data are a significant raw material of the 21st century. To exploit its value, a proper infrastructure that makes data FAIR – Findable, Accessible, Interoperable, and Re-purposable – is a must. For the wider field of the (chemical) physics of condensed matter and materials science, FAIRmat sets out to make this happen. Such data infrastructure enables extensive data sharing and collaborations in data-driven sciences, including artificial intelligence, and progresses basic science and engineering. FAIRmat engages with scientists across generations to advance careers, to promote materials science and innovations, and it reaches out to industry and society.

WHY TO ENGAGE IN FAIRmat?

The field of physics is extremely rich and complex, covering elementary particles, atoms, molecules, condensed and soft matter, chemical physics, surfaces, (quantum) optics, plasmas, astro-particles, and astronomy. As such, concepts and measuring techniques, working style, and research data are extremely diverse and heterogenous. Thus, **building a research infrastructure in a way that helps advancing basic and applied science rather than being a burden for the researchers** is an enormous challenge and can only be addressed and realized by the active scientists, from the bachelor level to professors and researchers in industry.

For sure, there must be several initiatives to tackle the diversity of the entire field, as even covering part of the effort in the required detail may appear like a hopeless endeavor. Indeed, in the overall NFDI concept, about three physics-related consortia, representing the extremely different areas, appear well-suited and conceivable. If and only if this process is performed as a bottom-up approach, driven by **experts in the respective field and embracing the community**, it is possible to succeed.

FAIRmat will realize this concept for materials science and related research fields. FAIRmat will be **inclusive**, making sure that nobody will be left behind. Importantly, FAIRmat is represented by the key scientists of the field, and it is **built on deep knowledge and already demonstrated experience**. The latter also includes outreach to and establishing synergies with other consortia, e.g. from chemistry or engineering. We are also discussing with the physics groups from other fields that were not yet ready

and did not submit a proposal in the first round. In short, FAIRmat comprises 7 *Areas*, listed below with the respective spokespersons:

AREA A: Synthesis - Martin Albrecht (Leibniz Institute for Crystal Growth) & Claudia Felser (Max Planck Institute for Chemical Physics of Solids)

AREA B: Experimental Materials Science - Mark Greiner (Max Planck Institute for Chemical Energy Conversion) & Christoph Koch (Humboldt-Universität zu Berlin)

AREA C: Computational Materials Science - Matthias Scheffler (Fritz Haber Institute Berlin) & Kurt Kremer / Tristan Bereau (Max Planck Institute for Polymer Research)

AREA D: Digital Infrastructure - Hans-Joachim Bungartz (Technical University of Munich) & Wolfgang Nagel (Centre for Information Services and High-Performance Computing, TU Dresden)

AREA E: Use Case Demonstrators - Christof Wöll (Karlsruhe Institute of Technology) & Axel Groß (Ulm University)

AREA F: User Support, Training & Outreach - Matthias Scheffler (FAIR-DI e.V.) & Martin Aeschlimann (TU Kaiserslautern)

AREA G: Administration & Coordination - Claudia Draxl & Matthias Scheffler (both FAIR-DI e.V.)

Each Area is subdivided in several *Tasks*; for more information, see www.fair-di.eu/fairmat/structure.

Figure 1: Christof Wöll (KIT and FAIR-DI), Markus Scheidgen (HU Berlin), Matthias Scheffler (FHI Berlin and FAIR-DI), Claudia Draxl (HU Berlin and FAIR-DI), and Mark Greiner (MPI for Chemical Energy Conversion) at the NFDI evaluation, Dec. 5-6, 2020.

STATUS OF THE FAIRmat PROPOSAL

FAIRmat submitted its proposal for being a consortium within the NFDI on 15 October 2019; the evaluation (*NFDI Colloquium Physics*) took place on December 5-6 in Königswinter near Bonn. FAIRmat was reviewed together with the other physics-related consortia, i.e. AstroNFDI (astronomy, astrophysics, and astroparticle physics), PAHN-PaN (particle, astroparticle, hadron, and nuclear physics), and DAPHNE (photon and neutron experiments). Overall, we experienced a very positive atmosphere and constructive questions. At the end of January, we received an excellent report, highlighting the significance of FAIRmat for the success of the NFDI and its leading role in an international context:

„*Das Konsortium wird als wesentlich für den Erfolg der NFDI eingestuft. Gleichfalls würde das Vorhaben mit einer international anerkannten Dateninfrastruktur für die naturwissenschaftlichen Materialwissenschaften aus Sicht der Begutachtenden eine Vorreiterrolle im internationalen Kontext übernehmen können.*"

The funding decision of the first round will be announced on June 26 by the GWK (Gemeinsame Wissenschaftskonferenz). We hope that FAIRmat will soon be able to start its work with full speed such to keep the engagement and enthusiasm of our members.

Over the past year, FAIRmat has created an euphoric spirit in the community. With NOMAD (www.nomad-coe.eu), it started out with computational materials-science (see below). Now, by reaching out to materials synthesis and experimental science, FAIRmat is covering the full complexity of the field. Being well integrated into the Condensed Matter Section of the German Physical Society, the MPG (e.g. Big-Data Network, CPTS), a large number of universities and institutes as well as various international activities (for example RDA, GO FAIR, EOSC), FAIRmat will quickly put its planning into practice, once a hopefully positive funding decision will have been made.

INTERNATIONAL AFFAIRS

FAIRmat and FAIR-DI are proud that the world's re-nowned experts in materials-data science and infra-structure have agreed to serve on their *International Advisory Board*. Their first members are:

Kristin Persson, director of The Materials Project (USA);

James Warren, director of the NIST Materials Genome Program (USA);

Isao Tanaka, a leader in Japan's big data activities, including high-through-put screening and AI (JP);

Barend Mons, co-leader of GO FAIR and president of CODATA (NL);

Kimmo Koski, co-initiator of EUDAT, managing director of CSC - IT Center for Science (FI).

Figure 2: Members of the International Advisory Board

NOMAD

Part of FAIRmat's **AREA C** is the NOMAD infrastructure (www.nomad-coe.eu). The *NOMAD Repository*, going online 2014 and being further developed as a European *Center of Excellence (CoE) in Computing Applications,* has established itself as the biggest data store of computational materials science worldwide. More than *100 million calculations* with billions of datasets have been uploaded at the NOMAD Repository. These were created by billions of CPU hours invested at high-performance computer centers worldwide.

The NOMAD CoE has now secured funding until fall 2023 as a CoE in High-Performance Computing (HPC). It is complementary to FAIRmat as it is focused on advancing scientific software from *ab initio* electronic-structure theory towards exascale computing and extreme-scale data. FAIRmat expects significant synergies with the NOMAD CoE also via the Barcelona Supercomputer Center (BSC) and the Finnish IT Center for Science (CSC) - two HPC centers where early 2021 the first pre-exascale computers will be installed in Europe.

RECENT EVENTS

From July 8 to 12, 2019, leading computational and experimental materials scientists got together with experts in data science at the *NOMAD/FAIR-DI Workshop on Shared Metadata and Data Formats for Big-Data Driven Materials Science* (for the outcome, see below).

From September 9 to 13, 2019, *Big Data Summer – A Summer School of the BiGmax Network* took place in Platja d'Aro, Spain. The school was predominantly targeted towards PhD students and young postdocs. The 15 invited speakers addressed important background and recent advances in data-driven materials science. The topics covered a wide spectrum to demonstrate the challenges and potential that research data offer.

James Warren (director of the materials genome program at NIST, USA) visited FAIR-DI and presented a lecture on *The Materials Genome Initiative and Artificial Intelligence* on January 30, 2020.

On January 20, 2020, the colloquium *FEST* (frontiers in electronic-structure theory) *with FAIRness* was held in honor of Claudia Draxl's birthday. Plenary speakers were Kristin Persson (Berkeley National Lab), Lucia Reining, this year's winner of the Gentner-Kastler-Prize of the DPG (ETSF, CNRS, École Polytechnique), Norbert Koch (HU Berlin), and Matthias Scheffler (FHI Berlin).

Figure 3: Kristin Persson, Norbert Koch, Lucia Reining, and Matthias Scheffler talking at the FEST with FAIRness, Berlin, January 20, 2020

Unfortunately, the **DPG Frühjahrstagung** of the Condensed-Matter Section (SKM) in Dresden had to be canceled. FAIRmat was significantly involved in three symposia:

- SYBD (Symposium *Big Data Driven Materials Science*; encompassing all SKM divisions): organized by J. Neugebauer and K. Kremer & M. Scheffler (both FAIRmat); invited talk by C. Draxl (FAIRmat)

- PSV III (*Impact of the upcoming 'Nationale Forschungsdaten Infrastruktur NFDI' on the SKM Community*): organized by E. Runge (SKM Chair and FAIRmat) and Young DPG; invited talk by M. Scheffler (FAIRmat)

- O 62 / CPP 66 (Focus Session: *Big Data in Aquisition in ARPES*): organized by R. Ernstorfer, M. Hartelt & M. Aeschlimann (all FAIRmat); invited talk by C. Draxl

These events reflect the enthusiasm with which the FAIRmat community is committed to the topic of data infrastructure and data-driven research. We are working on replacements of these symposia to be held in the near future.

OUTREACH AND EDUCATION

A **lecture series** on **Data-driven Materials Science** will be held at the Humboldt-Universität zu Berlin in collaboration with the Max-Planck Graduate Center for Quantum Materials. The first edition will start April 6, 2020. The lectures will be video-recorded and made available through the FAIRmat website.

FAIRmat members are currently discussing how to implement appropriate **curricula** to account for the challenges, chances, and the **interdisciplinary character of the NFDI** and the **future research** it will enable.

METADATA – A TRUE COMMUNITY EFFORT

In early 2016, the developers of the leading electronic-structure codes and the coordinators of the world-wide largest materials-science databases gathered at CECAM (a science and conference center in Lausanne) to discuss common principles for the standardization of the stored materials-science data.

The outcome of the workshop was summarized in the paper **Towards efficient data exchange and sharing for big-data driven materials science: metadata and data formats**, published in npj Computational Materials ([doi:10.1038/s41524-017-0048-5](doi:10.1038/s41524-017-0048-5)).

The crucial aspect here was the introduction of a hierarchical and extensible metadata schema, the NOMAD MetaInfo. It serves the annotation of all data contained in electronic-structure codes to allow for finding, accessing, and reusing the data.

At the **NOMAD/FAIR-DI workshop** on Shared Metadata and Data Formats for Big-Data Driven Materials Science (July 2019, HU Berlin) the next step of shared materials-science data was discussed, motivated by the need of **enabling interoperability of data across all computational and experimental sub-fields of materials science**. A key concept for adding semantic content to the annotations and

tracking provenance (including complex computational or experimental workflows) are ontologies.

All these aspects, their relation to the FAIR principles, and their integration into the NOMAD MetaInfo will be described in an upcoming community publication.

INVITED AND PLENARY TALKS

Since the launch of the NFDI call, the Area and Task Leaders of FAIRmat have given **well above 50 invited, keynote, and plenary talks** on FAIRness and the challenges and chances of building a Research Data Infrastructure (see figs. 4 and 5 and a constantly updated list on www.fair-di.eu/fairmat/talks).

Figure 4: Claudia Draxl at her plenary talk at the ICMS 2019, Jeju, Korea, November 3-6, 2019.
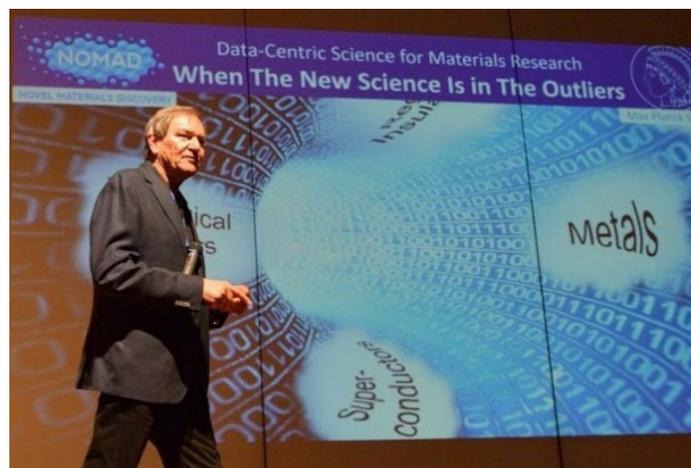

Figure 5: Matthias Scheffler at his keynote lecture at the Materials Research Meeting 2019, Yokohama, Japan, December 2019.

UPCOMING EVENTS

The **FAIR-DI Conference** on a **FAIR Data Infrastructure for Materials Genomics** will take place from June 3 to 5 at the Harnack House Berlin.

We were able to get the **top leaders of the field** as plenary speakers: Barend Mons (Leiden University Medical Center, NL, fig. 2), Claudia Draxl (HU Berlin & FAIR-DI e.V., DE, fig. 4), the Director of the NFDI, York Sure-Vetter (KIT, NFDI director, DE, fig. 6), James Warren (NIST Materials Genome Program, US, fig. 2), and Tong-Yi Zhang (Materials Genome Institute of Shanghai University, CN, fig. 7). This list is complemented by outstanding speakers for invited talks.

Figure 6: York Sure-Vetter

Figure 7: Tong-Yi Zhang

Application is possible via the conference website: https://th.fhi-berlin.mpg.de/meetings/fairdi2020/.

SOCIAL MEDIA

FAIRmat and FAIR-DI are on Twitter. Follow us @FAIRmat_NFDI and @FAIRDI_eV. NOMAD's Twitter is @NoMaDCoE.

IMPRESSUM

FAIRmat - FAIR Data Infrastructure for Materials Science, and Related Research Fields.

Spokesperson Claudia Draxl
FAIR-DI e.V. and Humboldt-Universität zu Berlin
Zum Großen Windkanal 6
12489 Berlin, Germany
fairmat@fairdi.eu
www.fair-di.eu/fairmat/consortium