

Introducing a FAIR RDM infrastructure for electron microscopy and other materials science data

Prof. Christoph Koch¹, Dr. Markus Kühbach¹, Sherjeel Shabih¹, Dr. Sandor Brockhauser¹, Prof. Dr. Erdmann Spiecker², Markus Scheidgen¹, Lauri Himanen¹, Ádám Fekete¹, Dr. José A. Márquez¹, Prof. Dr. Heiko Weber³, Prof. Dr. Claudia Draxl¹

¹Department of Physics and CSMB at Humboldt-Universität zu Berlin, Berlin, Germany, ²Department of Materials Science and Engineering, FAU Friedrich-Alexander-Universität, Erlangen, Germany, ³Institute of Condensed Matter Physics, FAU Friedrich-Alexander-Universität, Erlangen, Germany

Digitization and an increase in complexity and price of electron microscopy hardware and characterization techniques, as well as the maturation of machine learning tools to extract patterns from large amounts of very diverse (annotated) data, promise to accelerate materials development by synergistically combining research data from many sources. While some labs have started uploading their (raw) research data to data repositories, this is only a first but insufficient step to realize the above-mentioned potential, as such repositories are typically either specific to a very particular technique or agnostic to much of the domain-specific content of the uploaded data [1,2]. In both cases the research data cannot be easily compared and integrated with experimental data from other sources or numerical predictions, and certainly not without significant human effort,. Therefore, working towards an interoperable knowledge representation for experiments and computer simulations [3-6] is the main motivation for implementing FAIR research data management. This highlights the need for tools for information extraction and semantic mapping. Fundamental to these tools' effectiveness is the creation of thorough and transparent documentation. This needs to be made more complete, shared openly, and should benefit from activities where representatives of the communities agree on defining and using standardized knowledge representations.

We will report on recent progress by the FAIRmat NFDI consortium [7] in extending NOMAD, the world's largest data base for ab-initio computational materials data, to also host experimental research data on the synthesis and characterization of materials in a machine-accessible manner, i.e. annotated with well-defined and interoperable metadata that establish links between related (experimental and computational) quantities [8-10]. We will report on our work on developing a comprehensive data schema for electron microscopy and related techniques, and the corresponding software tools for data converting, visualizing, and online-processing. We have integrated these tools as customizations into NOMAD Oasis to offer a locally-installable version of the NOMAD research data management system to complement its note

keeping, file format parsing, cloud-based domain-specific data analyses, and information retrieval capabilities.

Keywords:

FAIR data management, database

Reference:

- [1] <https://www.re3data.org>
- [2] <https://explore.openaire.eu>
- [3] M. D. Wilkinson et al., (2016), <https://doi.org/10.1038/sdata.2016.18>
- [4] A. Jacobsen et al., (2020), https://doi.org/10.1162/dint_r_00024
- [5] M. Barker et al., (2022), <https://doi.org/10.1038/s41597-022-01710-x>
- [6] M. Scheffler et al., (2022), <https://doi.org/10.1038/s41586-022-04501-x>
- [7] <https://www.fairmat-nfdi.eu/fairmat>
- [8] M. Scheidgen et al., (2023), <https://doi.org/10.21105/joss.05388>
- [9] <https://github.com/FAIRmat-NFDI/>
- [10] <https://gitlab.mpcdf.mpg.de/nomad-lab/nomad-FAIR>