

Stepping Stones towards the Fourth Paradigm of Materials Science

Claudia Draxl

Physics Department and IRIS Adlershof, Humboldt-Universität zu Berlin
and Fritz Haber Institute of the Max Planck Society, Berlin

Novel approaches of Artificial Intelligence (AI) can find patterns and correlations in data that cannot be obtained from individual calculations or experiments and not even from high-throughput studies. In fact, data-driven research is adding a new research paradigm to the scientific landscape. For a real breakthrough, *Open Data* and sharing, as well as an efficient data infrastructure is key [1]. In other words, for shaping this forth paradigm and contributing to the development or discovery of improved and novel materials, data must be what is now called FAIR - Findable, Accessible, Interoperable, and Reusable [2].

The NOMAD Laboratory [3,4] is a living example for such infrastructure in computational materials science, comprising the NOMAD *Repository* (raw data) and its *Archive* (normalized, i.e. code-independent data). The NOMAD *Encyclopedia* is a web-based public platform that visualizes the results of this vast amount of calculations. The NOMAD Analytics Toolkit provides a collection of examples and tools to demonstrate how materials data can be turned into knowledge by AI approaches (e.g. [5,6]).

I will give a guided tour through this data lab and discuss the challenges [7] ahead of us to exploit materials data which are a remarkable raw material of the 21st century.

References:

- [1] C. Draxl and M. Scheffler, *Big-Data-Driven Materials Science and its FAIR Data Infrastructure*, Invited Perspective in Handbook Andreoni W., Yip S. (eds) Handbook of Materials Modeling. Springer, Cham (2019).
- [2] M. D. Wilkinson et al., The FAIR Guiding Principles for scientific data management and stewardship, *Sci Data* 3, 160018 (2016).
- [3] The NOMAD Laboratory, <https://nomad-coe.eu/>, including metadata and various software for parsing, analysis, and visualization.
- [4] C. Draxl and M. Scheffler, NOMAD: The FAIR Concept for Big-Data-Driven Materials Science, *MRS Bulletin* 43, 676 (2018).
- [5] L. M. Ghiringhelli, J. Vybiral, S. V. Levchenko, C. Draxl, and M. Scheffler, *Big Data of Materials Science - Critical Role of the Descriptor*, *Phys. Rev. Lett.* 114, 105503 (2015).
- [6] L. M. Ghiringhelli, J. Vybiral, E. Ahmetcik, R. Ouyang, S. V. Levchenko, C. Draxl, and M. Scheffler, *Learning physical descriptors for materials science by compressed sensing*, *New J. Phys.* 19, 023017 (2017).
- [7] C. Draxl and M. Scheffler, *The NOMAD Laboratory: From Data Sharing to Artificial Intelligence*, *J. Phys. Mater.* 2, 036001 (2019).