On the Search for Novel Materials:
Insight and Discovery through Sharing of Big Data

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

On the steady search for advanced materials with tailored properties and novel functions, high-throughput screening is a new branch of materials research. For successfully exploring the chemical compound space from a computational point of view, two aspects are crucial. These are reliable methodologies to accurately describe all relevant properties for all materials on the same footing, and new concepts for getting insight into the materials data that are produced since many years with an exponential growth rate.

In this talk, I will discuss our concepts for tackling big data of materials science. It is not an issue of boosting more high-throughput calculations but it is about the question: *How to exploit the wealth of information, inherently inside the materials data which promises unprecedented insight?*

I will first introduce the *NoMaD Repository* [1], which was established to promote the idea of open access and sharing of materials data. As open access implies that data can be used by anyone, large collections of materials data opens an avenue for using and developing tools that the present (computational-)materials community does not even know. The latter will be realized in the Novel Materials Discovery (NOMAD) Center of Excellene (CoE) [2], the main aims of which are the creation of a *Materials Encyclopedia* and the development of big-data analytics tools for materials science. Finally, I will demonstrate an example how statistical-learning approaches based on domain-specific knowledge can indeed lead to new scientific insight [3].

[1] The Novel Materials Discovery (NoMaD) Repository: https://nomad-repository.eu

[2] NOMAD Center of Excellence, funded by the EU within HORIZON2020: http://nomad-CoE.eu

[3] L. 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).