

## MM 51: Topical session: Data driven materials design - databases

Time: Wednesday 17:15–18:00

Location: BAR 205

MM 51.1 Wed 17:15 BAR 205

**The NOMAD (Novel Materials Discovery) Laboratory: Concepts, Challenges, and Results** [∗] — ●PASQUALE PAVONE<sup>1</sup>, GEORG HUHS<sup>1,3</sup>, LUCA GHIRINGHELLI<sup>2</sup>, CLAUDIA DRAXL<sup>1,2</sup>, and MATTHIAS SCHEFFLER<sup>2</sup> — <sup>1</sup>Humboldt-Universität zu Berlin, Physics Department and IRIS Adlershof, Germany — <sup>2</sup>Fritz-Haber-Institut der MPG, Berlin, Germany — <sup>3</sup>Barcelona Supercomputing Center, Spain

The NOMAD (Novel Materials Discovery) Laboratory Center of Excellence (CoE), <https://NOMAD-CoE.eu>, is a community-driven activity with the mission to serve the whole field of materials science and engineering. It tackles the issues of Big Data in materials science, starting from the NoMaD Repository [1], which by now contains results from more than 18 million DFT total-energy calculations. This corresponds to several billion CPU-core hours spent on high-performance computers worldwide. With these and more and more incoming data, we build a Materials Encyclopedia, to provide a user-friendly access to all these results, also making use of Advanced Graphics. Novel Big-Data Analytics tools are developed for finding trends, identifying outliers, and predicting new materials with tailored properties. Thus, this talk describes the basic concepts and goals of the NOMAD CoE in order to stimulate a discussion and suggestions on how to proceed further. This work received funding from “The Novel Materials Discovery (NOMAD) Laboratory”, a European Center of Excellence.

[∗] Collaboration with the entire NOMAD team: <https://nomad-coe.eu>  
[1] <https://NOMAD-Repository.eu>

MM 51.2 Wed 17:30 BAR 205

**The NOMAD Analytics Toolkit: Interactive Big-Data Driven Materials Science over the Web** — ●LUCA M. GHIRINGHELLI, FAWZI MOHAMED, ANKIT KARIRYAA, ANGELO ZILETTI, CHRISTIAN CARBOGNO, and MATTHIAS SCHEFFLER — Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin, Germany [‡]

Big-data analytics opens new routes towards scientific insight by offering analysis tools that can reveal patterns, trends, and causal relationships so-far hidden in the data. The Novel Materials Discovery (NOMAD) Repository (<https://nomad-repository.eu>) stores millions of open-access materials-science calculations obtained with dozens of different codes. In order to perform analytics on this data, two further steps are necessary: i) the raw, code-specific inputs/outputs of the calculations need to be converted into a stan-

dardized format that uses one convention for, e.g., units, zero base lines, and file formats. We present a flexible and extensible meta-data infrastructure (<https://metainfo.nomad-coe.eu>), implemented for storing the data in an intuitive, code-independent, representation. ii) the data needs to be easily and publicly searchable, retrievable, and analyzable. With the NOMAD analytics toolkit (<https://analytics-toolkit.nomad-coe.eu/>), we present an interactive web-interface that allows everybody, without need to install any software, to query and analyze the data. We demonstrate the toolkit with an analysis of oxide semiconductors data, looking for a structure-property relationship with statistical methods.

[‡] Collaboration with the full NOMAD team: <https://nomad-coe.eu>.

MM 51.3 Wed 17:45 BAR 205

**The NOMAD Repository - a key service for the computational-materials science community** — ●JUNGHOO SHIN<sup>1,2</sup>, THOMAS ZASTROW<sup>3</sup>, LORENZO PARDINI<sup>1</sup>, STEFAN HEINZEL<sup>3</sup>, MATTHIAS SCHEFFLER<sup>2</sup>, and CLAUDIA DRAXL<sup>1,2</sup> — <sup>1</sup>Humboldt-Universität zu Berlin, Physics Department and IRIS Adlershof, Germany — <sup>2</sup>Fritz-Haber-Institut der MPG, Berlin, Germany — <sup>3</sup>Max Planck Computing and Data Facility, Garching, Germany

The NoMaD Repository, <http://nomad-repository.eu/>, is a community effort, offering to store and share input and output files of all popular electronic-structure codes. The uploaded files are saved for at least 10 years. Thus, uploaders and their groups can get back what they have done in the past, and also others can use the data for their own scientific work. The NoMaD Repository offers DOIs free of charge which makes the scientific data are citable. It is the only repository in materials science recommended by Nature Scientific Data. Currently, the NoMaD Repository contains more than 3 million entries, which corresponds to more than 18 million total-energy calculations and several billion CPU core hours spent on various high-performance computers worldwide. Recently, the open-access data have been parsed and normalized by the NOMAD Laboratory, <https://nomad-coe.eu/>. Its Archive contains normalized, (largely) code-independent data, which is an important prerequisite for data-driven materials science. This work received funding from “The Novel Materials Discovery (NOMAD) Laboratory”, a European Center of Excellence.

(\*\*) Collaboration with the entire team: <http://nomad-repository.eu>