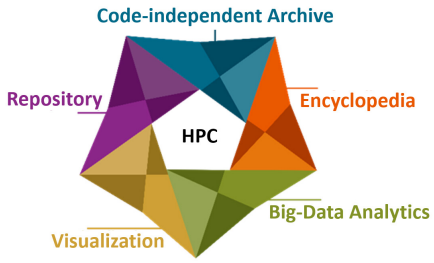
The NOMAD (Novel Materials Discovery) Laboratory
A European Centre of Excellence

IN THIS ISSUE - November 2018



The Novel Materials Discovery (NOMAD) Laboratory CoE maintains the largest Repository of input and output files of all important computational materials science codes. From its open access data, it offers several Big-Data Services to advance materials science and engineering.

Loss of Influential NOMAD Leaders

It is with great sadness that we mark the passing of two influential NOMAD leaders, who will be greatly missed: Alessandro De Vita and Stefan Heinzel.

Alessandro De Vita (King's College London) tragically passed away on 02 October 2018. With him, we lost a co-founder and director of the Thomas Young Centre (https://www.thomasyoungcentre.org/), the London center for theory and simulations of materials, a Principal Investigator and one of the driving forces behind NOMAD. Sandro was a forerunner of artificial intelligence applied to materials science. His "learn on the fly" approach practically opened a new field. We will miss his profound scientific insight, sharp judgement and caustic wit uniquely mixed with Mediterranean warmth and cheerfulness.

Stefan Heinzel passed away on 28 August 2018. We lost the competent leader of the Max Planck Computing and Data Facility, an influential and reliable NOMAD and BiGmax Principal Investigator (https://www.bigmax.mpg.de/). Stefan was a role model of a leader, and it was most impressive how he and his team were running the MPCDF. He was such a wonderful and helpful person and so important for the Max Planck Society. We will miss his friendly and straight personality, his skills, his competent judgement, his reliability, and his sense of obligation.

NOMAD: The FAIR Concept for Big-Data-Driven Materials Science

Claudia Draxl presented NOMAD's FAIR concept at a webinar hosted by the Materials Research Society on 26 September 2018 (https://mrs.digitellinc.com/mrs/sessions/31555/view). The presentation highlighted how the computational materials science field has embraced Open Science, changing its scientific culture towards comprehensive and Findable, Accessible, Interoperable, and Reusable data. FAIR data open exciting new avenues for Big-Data-driven science, the 4th Paradigm of Materials Science and Engineering (see photo background, p. 3 "Making the Data Revolution Happen"). The webinar related to the September issue of the Materials Research Bulletin that focused on Data-Centric Science for Materials Innovation, with NOMAD's contribution now online: C. Draxl and M. Scheffler, NOMAD: The FAIR Concept for Big-Data-Driven Materials Science, MRS Bulletin (2018) 43:676, also http://arxiv.org/abs/1805.05039.

FAIR Data Infrastructure for Physics, Chemistry, Materials Science, and Astronomy e.V. (FAIRDI e.V.)

In a similar spirit to that which built up and advanced the internet, the time is more than ripe for building an infrastructure that enables efficient collection and sharing of data. Without such a reliable infrastructure, the envisioned success of Big-Data-driven materials science will be hampered. Therefore, researchers from Germany (Fritz Haber Institute, Humboldt University, Karlsruhe Institute of Technology, Fraunhofer-Institut für Werkstoffmechanik, Forschungszentrum Jülich), and the Netherlands (the NWO Institute ASTRON, Radboud University Nijmegen, University of Twente) have been enthusiastically discussing for several months how to join forces in order to take the scientific community to the next level. In early September 2018, the founding meeting of the non-profit association FAIR Data Infrastructure for Physics, Chemistry, Materials Science, and Astronomy – in short FAIRDI – has finally taken place in Amsterdam.

The association (https://fairdi.eu) will support science and research in various areas: Pillar A - Computational materials science, Pillar B - Experimental materials science, Pillar C - Biophysical and soft-matter simulations, Pillar D - Astrophysics and space-situational awareness, and Pillar E - User management, intellectual property rights, cyber security (horizontal).

The specific goals are to strongly support extensive sharing of scientific data to advance research in science and engineering (thereby also preventing scientific misconduct); to host raw and normalized data, such that data from different studies can be compared and used for other purposes than those initially intended when the data were created; to build an infrastructure to make Big Data easily available to other computer centers and research labs in academia and industry; and to make data ready for analysis by methods from artificial intelligence.

We are confident that FAIRDI will grow into a European, and ultimately international, network!

NOMAD Summer 2018

Four keynote speakers, 26 speakers and tutors from the NOMAD team, and 34 participants made *NOMAD Summer* 2018 a resounding success. Held at the CECAM Headquarters (Lausanne, 24 - 27 September 2018) overlooking Lake Geneva, the school introduced researchers from academia and industry to state-of-the-art methods and practical tools in Big-Data-driven materials science developed within NOMAD. The program included keynote addresses, lectures and hands-on tutorials, but also allowed for numerous and interesting discussions with participants, including representatives of world-leading companies. The greatest interest was in the application of the various tools developed by NOMAD to industrially-relevant problems and how quality can be assured, as well as scientific reproducibility.

Intensive discussions with industry representatives at NOMAD Summer 2018 - From left: Claudia Draxl (NOMAD Encyclopedia), Matthias Scheffler (NOMAD Coordinator), Deepak Jain (Tata Consultancy Services Limited - Tata Research Development & Design Centre, India) and Markus Rampp (NOMAD Advanced Visualization).

Following the popular *NOMAD Summer 2017*, we had more interested participants than space available - so, if you were unable to join the school, videos are available of all the talks on the program page (http://meetings.nomad-coe.eu/nomad-summer-2018/index.php?n=Meeting.Program).

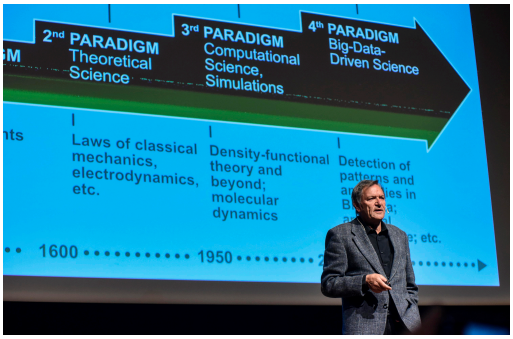
A movie about the *NOMAD Big-Data Analytics Toolkit* (https://th.fhi-berlin.mpg.de/movies/?m=nomad-bat) is also now available and gives a short introduction on its purposes and main features, together with an overview on what tools do exist and what possibilities they offer.

Making the Data Revolution Happen

The Global Internet of Things Summit (GIoTS) conference brought together researchers and industry communities active in the IoT domain from all over the world and was endorsed by the world's largest technical professional organization for the advancement of technology, the Institute of Electrical and Electronics Engineers (IEEE). NOMAD Coordinator Matthias Scheffler gave the ONLY Plenary Keynote Speech, "Making the Data Revolution Happen - Turning Billions of Data from Computational Materials Science into Knowledge by Artificial Intelligence" at the GIoTS 2018 Opening Plenary Session on 05 June 2018.

In the audience were top officials, including the Deputy Minister for Technology, Innovation and Competitiveness of the Government of the Basque Country and a Delegate Councilor for Urban Planning of the City Council of Bilbao. Overall, 800 participants attended the IoT summit.

Matthias also spoke at the workshop "Making the Data Revolution Happen: Machine Learning and Big Data - Will it help transforming Data into Knowledge?" on 04 June 2018. The goal of the workshop was to discuss new Big Data processing and analytics approaches, which will be required to create useful knowledge from the massive data streams generated by billions of smart IoT devices.



Plenary Keynote Speech by Matthias Scheffler highlights the 4th Paradigm in Materials Science, where Big-Data is driving novel materials discovery.

Big Data vs. Fast Computation - Is HPC Facing a Game Change? - Panel Discussion at PASC 2018

PASC2018 Panellists Eng Lim Goh (Hewlett-Packard), Torsten Schwede (University of Basel), Matthias Scheffler (Fritz Haber Institute), and Nuria Lopez (University of Barcelona).

The Platform for Advanced Scientific Computing (PASC) conference was held from 02 - 04 July 2018 in Basel, Switzerland. PASC is an international platform for the exchange of competences in scientific computing and computational science, with a strong focus on methods, tools, algorithms, application challenges, and novel techniques and usage of high performance computing. The focus of this year's PASC conference was *Fast and Big Data, Fast and Big Computation*.

On Tuesday, 03 July 2018, NOMAD Coordinator Matthias Scheffler was one of four experts invited to discuss the critical challenges arising in the next 5-10 years for HPC application areas at the Panel Discussion "Big Data vs. Fast Computation – Is HPC Facing a Game Change?". This Panel Discussion was initiated and moderated by Florina Ciorba and Erik Lindahl and can be found on-line on YouTube (https://www.youtube.com/watch?v=mTqzCvm0G5c).

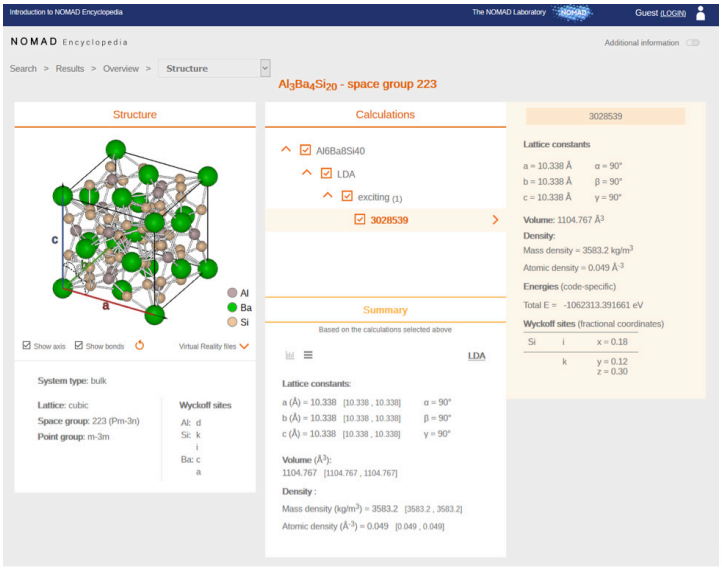
New Release of *NOMAD Encyclopedia*

The *NOMAD Encyclopedia* has recently seen a major update by BSC, HUB and the entire development team. Its data content has grown from half a million to 1.3 million different materials. The graphical presentation of its structure module has been equipped with more details on the atomic structure and its classification. The presentation of methodology details has become more user-friendly and was extended with filtering functionality. Several of these extensions were initiated based on feedback and needs from *NOMAD Encyclopedia* users, predominantly industrial researchers.

The *NOMAD Encyclopedia* and its API has raised interest among academic and industrial users for automated searches for materials and data processing. One example is a toolkit, created by Martin Kuban, a PhD student at HUB, which evaluates the similarity of materials in terms of their electronic structure. This approach was inspired by, but goes beyond, a fingerprinting scheme of the density-of-states (DOS), introduced by Isayev and coworkers (O. Isayev et al., Chem. Mater. (2015) 27:735). For the development process, the *NOMAD Encyclopedia* API was used to collect DOS data for about 280,000 materials. The *NOMAD Encyclopedia* developers team is currently exploring how to integrate such tools in the GUI.

Other novel features will soon be released. Most important, a new search interface for complex searches is in its final development phase: The new layout allows users to freely combine chemical formulae and elements when searching for materials with certain chemical composition. For example, searches like “all materials containing AlO_3 as well as any of the group 11 elements” will become possible. Additionally, a new query-technology behind this functionality provides much better performance.

Other major extensions concern the classification and presentation of surface and adsorbate systems, an interactive toolbox for the analysis of elastic constants and the combined representation of results from related calculations.



Representation of structural details in the NOMAD Encyclopedia, exemplified for the clathrate $Al_3Ba_4Si_{20}$. On the left, below the interactive ball-and-stick-model, the Wyckoff sites of inequivalent atoms are listed. The Si atoms in the given structure occupy two different Wyckoff sites; their free parameters, which might differ amongst different calculations, are provided on the right hand side.

NOMAD - a GO FAIR Implementation Network

NOMAD has become an Implementation Network of GO FAIR (https://www.go-fair.org/), a bottom-up international approach for the practical implementation of the European Open Science Cloud (EOSC) as part of a global Internet of FAIR Data & Services.

Pasquale Pavone represented NOMAD at the recent GO FAIR event “GO BUILD - GO TRAIN - GO CHANGE” in Berlin, Germany, 19 September 2018.

exciting and NOMAD - from Berlin to Shanghai

From 31 July to 09 August 2018, NOMAD's team at the Humboldt University Berlin hosted "HoW exciting! 2018" (http://exciting-code.org/how-exciting-2018), a hands-on tutorial on excitations in solids, followed by an international workshop with world-leading experts. The event included lectures from a number of NOMAD scientists, including Claudia Draxl, Andris Gulans, Patrick Rinke, Pasquale Pavone and Matria Troppenz. This year, the event was co-organized with the EU Training Network INFORM.

HoW exciting! 2018 hands-on tutorial participants.

Electronic-structure theory, excitations, and Big Data are also the key topics of the "Exciting Shanghai School" (http://icqms.shu.edu.cn/exciting_shanghai_2018/), that will take place 17 - 22 November 2018. It consists of keynote lectures on first-principles simulations in materials science and hands-on tutorials. This event is jointly hosted by the International Centre for Quantum and Molecular Structures (ICQMS) and the Materials and Genomics Institute (MGI) of the Shanghai University, Shanghai, China. Together with Wei Ren (Shanghai University, Chine) and Alessandro Stroppa (CNR-SPIN, Italy), NOMAD's Pasquale Pavone and Claudia Draxl are among the organizers.

2018 CAMD Summer School

The sixth Computational Atomic-scale Materials Design (CAMD) Summer School on Electronic Structure Theory and Materials Design (https://www.fysik.dtu.dk/english/Research/CAMD/Events/Summer-school-2018, supported by Psi-k) was held in August 2018 in the historical town of Helsingør north of Copenhagen. Twenty world-leading scientists, including NOMAD's Claudia Draxl and Kristian Thygesen, lectured to the 140 participating students on topics ranging from basic electronic structure theory and many-body perturbation theory to materials databases and machine leaning methods. In line with the NOMAD project, many lecturers emphasized the importance of data-driven approaches to materials discovery.

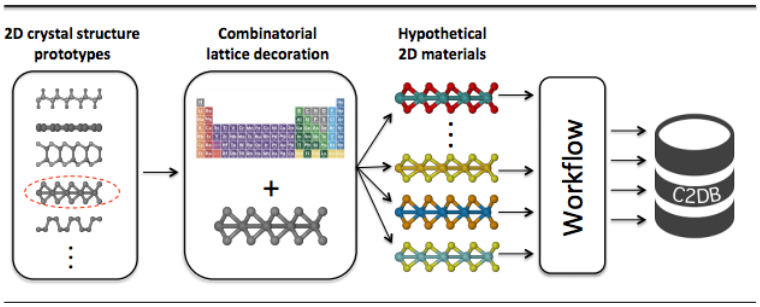
2018 CAMD Summer School participants.

The Computational 2D Materials Database (C2DB)

Over the past decade, atomically thin two-dimensional (2D) materials have made their way to the forefront of several research areas, including batteries, (electro-) catalysis, electronics, and photonics. This development was prompted by the intriguing and easily tunable properties of atomically thin crystals and has been fueled by the constant discovery of new 2D materials and the emergent concepts of lateral and vertical 2D heterostructures, which opens completely new possibilities for designing materials with tailored and superior properties.

In the September 2018 Psi-k Highlight (http://psi-k.net/highlights/), DTU-based researchers led by NOMAD PI Kristian Thygesen, introduced "The Computational 2D Materials Database (C2DB)". This comprehensive database organizes a variety of *ab initio* calculated properties (including structural, elastic, electronic, magnetic, dielectric, and optical properties) of around 3,000 2D materials. In the paper, the researchers describe the workflow behind the database, present an overview of the properties and materials currently available, and explore trends and correlations in the data. A large number of new, potentially synthesizable 2D materials with interesting properties are identified targeting applications within spintronics, (opto-) electronics, and plasmonics.

The C2DB offers a comprehensive and easily accessible overview of the rapidly expanding family of 2D materials and forms an ideal platform for computational modeling and design of new 2D materials and heterostructures. The database is fully open and can be browsed online (http://c2db.fysik.dtu.dk) or downloaded in its entirety, also from the *NOMAD Repository*.

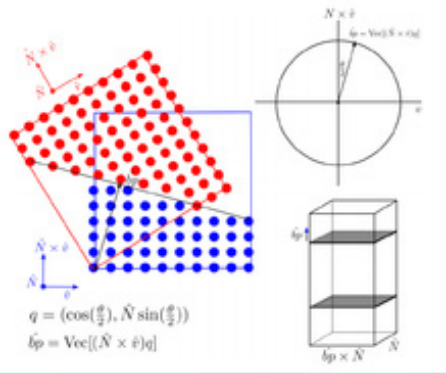


Materials in the C2DB are initially generated by decorating an experimentally known crystal structure prototype with atoms chosen from a (chemically reasonable) subset of the periodic table.

Open Access Grain Boundary Tools from NOMAD

The NOMAD teams at KCL and the University of Warwick have published new open access resources for the Imeall package, used to compute and index the atomistic properties of grain boundaries (Imeall: A computational framework for the calculation of the atomistic properties of grain boundaries, H. Lambert, A. Fekete, J.R. Kermode, and A. De Vita, Computer Physics Communications (2018) 232:256). The Imeall package provides a structured database for the storage of atomistic structures and their associated properties, equipped with a programmable application interface to interatomic potential calculators. The resources described are available online and can be extended by any user.

All computed atomistic properties are exposed via a web framework, providing open access to the grain boundary repository and the analytic tools suite (Code repository: github.com/kcl-tscm/imeall; Links to the full structure database, hosted on NOMAD servers: kcl-tscm.github.io/imeall/docs/build/index.html; Database: imeall.co.uk/). The server space is reserved for the next 10 years, ensuring mid- to long-term storage and public accessibility.



General coordinate system for determining tilt and twist boundary planes and orientations using quaternion algebra.

All NOMAD open access publications are available here: www.nomad-coe.eu/outreach-publications.

Hands-On DFT and Beyond Summer School

NOMAD Scientists traveled to Beijing this summer (30 July to 10 August 2018) to teach and tutor at another installment of the Summer School series "Hands-On DFT and Beyond" this year under the greater topic "Frontiers of Advanced Electronic Structure and Molecular Dynamics Methods" and generously funded by the Sino-German Center for Science Promotion, a joint effort by Germany's DFG and China's NSFC.

Topics ranged from the very basics of electronic-structure theory to their application for materials and molecules of real interest by means of approaches to sampling/structure search, thermodynamics, statistical mechanics, transport, excited states and spectroscopy, and finally also to Big-Data Analytics in computational materials science. In total, 68 students, 27 lecturers, and 14 tutors enjoyed the open and creative atmosphere at Peking University's School of Physics. The school was organized by Carsten Baldauf and Matthias Scheffler of Fritz Haber Institute in Berlin, Volker Blum from Duke University in Durham, Xinguo Ren from the University of Science, and Technology of China in Hefei, and in particular by Xinzheng Li and Hong Jiang from the Peking University.

"Hands-On DFT and Beyond" participants enjoyed both the lecture theatre and the local sights.

IMPRESS 2018

Interfacing Machine Learning and Experimental Methods for Surface Structures (IMPRESS) 2018 workshop (https://www.if.tugraz.at/impress/) was held 11 - 13 July 2018 at the Graz University of Technology, Austria. Co-sponsored by NOMAD, Psi-k, and CECAM, the workshop brought together experimentalists and theorists to tackle key challenges, including defects and grain boundaries, commensurability, thermodynamics versus kinetics, processing conditions, experimental error bars, multilayer and morphology, and polymorph identification and retrieval.

NOMAD at the University of Science and Technology China

The University of Science and Technology China (USTC) is one of the very best universities in science and engineering in China. USTC organized the Great Minds Forum on 10 October 2018, where the NOMAD coordinator, Matthias Scheffler, gave a 90-minute colloquium (see the announcement poster). He also discussed with leading scientists the installation of a *NOMAD Mirror* in China. The first steps to establish such a mirror have already been taken, and the open-access data of the *NOMAD Repository* have been transferred. The user management remains at the Max-Planck Computing and Data Facility (MPCDF) in Garching.

USTC poster announcing NOMAD Colloquium by Matthias Scheffler (size about 4.5m x 3.5m).

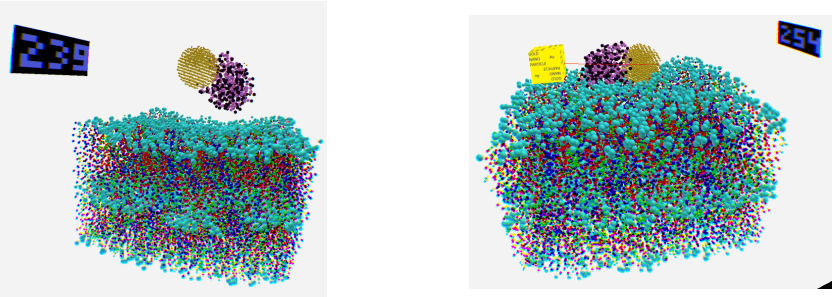
NOMAD Virtual Reality Tools Applied to Drug Discovery

NOMAD Virtual Reality (VR) tools can be used to experience more than just materials science data!

Working with Deepak Jain of TATA Consultancy Services (www.tcs.com), a Global Leader in IT Services, NOMAD's Rubén García-Hernández developed the VR visualization of Jain's new *in-silico* model of skin. In their tutorial (www.nomad-coe.eu/vrtutorial8), you can watch drugs pass through virtual skin.

This technology holds enormous potential for drug companies - new, better drugs that can be applied on the skin rather than injected can be discovered using computers instead of time-consuming, costly experiments with cells in laboratories or animal models.

There are many demos and videos of the exciting VR tools developed by NOMAD (www.nomad-coe.eu/VR-prototype) and the NOMAD VR User Manual is now available - have a look!



HTC Vive images of a drug molecule (left) passing through a virtual model of skin (right).

Institute for Pure & Applied Mathematics Long Program

The Institute for Pure & Applied Mathematics (IPAM, www.ipam.ucla.edu/) has organized a Long Program Science at Extreme Scales: Where Big Data Meets Large-Scale Computing (12 September - 14 December 2018) to bring together leading innovators and pioneers from applied mathematics (scientific computing, optimization, data analytics, statistics etc.), computer science (HPC, data engineering, data analytics, visualization, imaging etc.) and various applications areas to catalyze the new, emerging field linking HPC with Big Data. NOMAD's Claudia Draxl of Humboldt University is an organizer of this Long Program which consists of an opening day, tutorials, a series of workshops, and a culminating workshop at UCLA's Lake Arrowhead Conference Center. Claudia Draxl is also on the organizing committee of the related workshop on HPC and Data Science for Scientific Discovery, which took place 15 - 19 October 2018.

UCLA, Long Programme Venue.

During this event, Matthias Scheffler gave an invited talk on "When more data do not provide a better description" and Claudia Draxl gave an invited talk on "How FAIR are data repositories in materials science?", both on Friday, 19 October 2018. Videos are available here: http://www.ipam.ucla.edu/programs/workshops/workshop-ii-hpc-and-data-science-for-scientific-discovery/?tab=schedule).

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 676580. The materials presented and views expressed here are the responsibility of the author(s) only. The EU Commission takes no responsibility for any use made of the information set out.