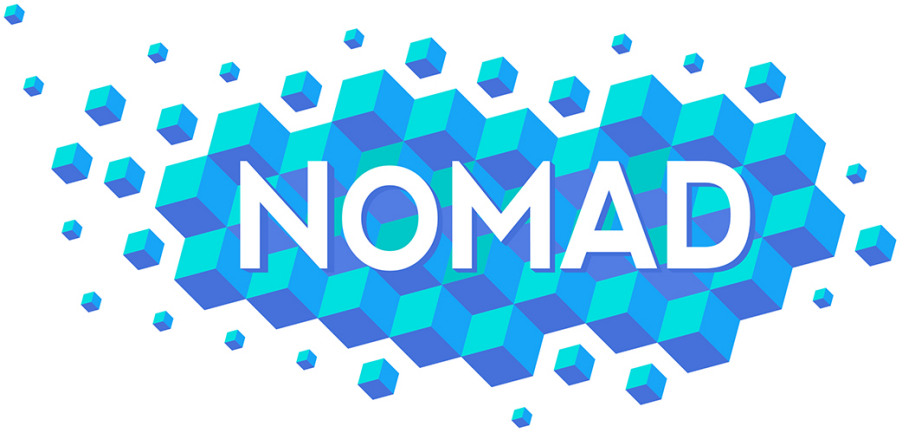

Title: 3rd NOMAD (Novel Materials Discovery) Industry Workshop

Location: Cumberland Lodge, Windsor, Berkshire, SL4 2HP

Sponsors: CECAM and Psi-K

Dates: February 5, 2018 to February 6, 2018

Organizers: Alessandro De Vita (King's College London), Angel Rubio (Max Planck Institute for the Structure and Dynamics of Matter, Hamburg, Germany), Matthias Scheffler (Fritz-Haber Institute, Berlin, Germany)

Summary

The objectives of NOMAD Centre of Excellence (CoE) include the creation of a materials encyclopaedia, the development of Big-Data analytics and advanced graphics tools for materials science and engineering. These goals are complementary with those of the other two CoEs supported by the European Commission and active in the field of CECAM activities (E-cam and Max). The NOMAD Researchers are currently creating a large, homogenized materials database, as well as the analytical tools and code developments necessary to extract information from it. We are confident that the available HPC infrastructure and the envisaged developments will contribute to the discovery of new scientific phenomena, novel devices, and advances in materials science and engineering. An essential corner stone of the project is the NOMAD Repository which contains the produced data in the form of input and output files of many high-quality calculations performed by fellow researchers working all over the world. The Repository is unique as it is not restricted to a few simulation programs but it accepts entries from all relevant codes. As of early January 2018, it contains results from more than 50 million total energy calculations, corresponding to more than 2 billion CPU-core hours used on high-performance computers all over the planet. Recently, there has been a very

significant growth of industrial interest (from large companies as well as SMEs) in computational materials science, motivated by the innovation potential of new materials and improved existing materials. However, the scientific complexity of the topic and the heterogeneous and fragmented nature of the research field make it difficult to translate research leadership into accomplished innovation. Thus, it is particularly important at this point in time to bring materials modelling closer to industrial/societal exploitation in real terms. We think that workshops with industry representatives are indispensable to gather such feedback.

Description and outcomes

This was the third (and last) of a series of three industry meetings organised annually by NOMAD to get together with industry representatives. The purpose of the meeting is to listen and gather the feedback of industry on their needs and plans concerning materials data, and to in- form/train them on data-analytic tool-usage. In addition, we share the recent developments NOMAD has carried out. At the end of each meeting, a commission made of NOMAD PI's and selected Industrial representative discuss the outcome of the meeting and plan in which direction NOMAD development should go in order to meet the industry needs. In particular, the meeting is structured such that in a first instance, invited industry representatives present and discuss the main activities carried out in their company. In second instance, speakers from each NOMAD work package present the most recent developments and features incorporated into the NOMAD's framework. Below we summarise what has been discussed for each work package.

Industry Networking:

NOMAD gathered very useful feedback from industry in its strive to determine how to make NOMAD useful for industry. Based on feedback from industry representatives and industrial advisory committee (IAC) members, several recommendations have been made some of which were already implemented. For example, an on-site usage of the platform has been provided for industrial users during the development. In order to ensure uptake of NOMAD tools and services by industry, the NOMAD team wishes to ensure that there are minimal barriers to use for industry users. Industry representatives have suggested the possibility of facilitating in-house usage of NOMAD to minimize potential intellectual property rights (IPR) issues. Furthermore, several groups within NOMAD are currently involved with developing case studies, some of which were showcased at the CECAM workshop to illustrate practical application of the NOMAD tools.

NOMAD Encyclopaedia:

NOMAD Encyclopaedia is accessible to various types of users (engineers, material scientists, non-experts) and provides general reviews and statistics via a WEB interface. Up until now, over 400,000 materials have been registered. As a result of the suggestions collected during the industry meetings, a series of new features have been introduced. Among others, this includes a newly developed user-friendly GUI, which enables users to perform property-based searches. Upon feedback from the

workshop, the NOMAD team will add more specific system types and methodologies to the already existing features.

NOMAD Big-Data Analytics Toolkit:

The objective of the NOMAD Big-Data Analytics Toolkit is to facilitate the design of new materials with desired properties; many of the tools apply machine-learning methods to data originating from computational high-throughput methods.
The structure of the NOMAD Big-Data Analytics Toolkit is a combination of a software framework and notebook representation of the actual analytic task.
Acting upon feedback from the workshop, the NOMAD team will keep extending the breadth of its analytic tools, in terms of both methodology and example cases.

NOMAD Advanced Graphics:

The innovative methods developed within this NOMAD work package allow both for local visualization running on researcher's workstations, as well as remote visualization tools running on HPC clusters. Local visualization tools also cover immersive virtual reality (VR) environments. Local tools are complemented by remote visualization tools, which allow for visualization of datasets too large for ordinary workstations, but with a lesser degree of interactivity. The general feedback gathered at the workshop pointed at establishing a closer integration between the Advanced Graphics and the NOMAD Encyclopaedia. After a previous workshop on the visualization tools took place in September 2017, a specific one for VR tools usage is now being planned (April 2018), open to the participation of industry representatives.

Outreach and connecting academia with Industry:

A special student-oriented tutorial on Machine-learning and VR based on data contained in the NOMAD repository relative to an industry case study was run during the workshop, which turned out to be much appreciated. Feedback was also gathered on the perceived exposure of materials modelling students to the practices and needs of industrial R&D, which turned out to be deemed largely insufficient by over 90% of the attending students and young researchers.

Further discussion on community needs

The two main concerns emerged in discussions with the workshop industry delegates addressed IP and Sustainability issues, while a considerable time was spent addressing issues related to data quality and usability. Furthermore, dissemination measures and how to increase the visibility of NOMAD were also addressed as one in need of attention. In this context, the case studies presented at the workshop to demonstrate NOMAD value have been initiated and are in continuous development, to help internal publicising of the value implied collaborating with the NOMAD team. In more detail, the emerging consensus was that:
IP: A top priority is to protect commercial data, which should be allowed not to leave commercial environments. NOMAD's tool and data should be able to be

downloaded and kept behind the industrial firewalls before performing any kind of data analytics on it. The local installation of data analytics tool also raises the issue of managing sensitive data and IP, while the suggestion that part of proprietary data might be volunteered to the public at large was received with mixed enthusiasm among our attending industrial partners.
Data and Usability: The wish for more tutorial services (training material/courses) and a user-friendly GUI interface to reduce the barrier to use the NOMAD tool was meanwhile clearly expressed and iterated. Searches for properties, rather than materials as well the integration of additional data on a wider range of material properties and experimental data etc. are features that, it was felt, greatly improve the usability of NOMAD. One of the main challenges of using NOMAD for industrial R&D remains how to combine in-house data with NOMAD's data for cross-analysis, as for this the in-house proprietary data must be similarly formatted, and must be of a similar quality to NOMAD. Delegates greatly appreciated NOMAD's effort to be visible, to continuously develop case studies, and to guarantee the safeguard of archived data.

Considering that one of the main goals of NOMAD is to be useful to the industry and work in close collaboration with them, the project is already successfully delivering several societal benefits:
- It is setting a standard on how to deal with large as well as small-sized datasets (for both quality and consistency).
- It enables efficient searches for a wide range of materials properties thus enabling the selection of the most promising materials for specific applications, as needed in industrial R&D. This involves significant saving time and economic resources, which will be beneficial for the industry. Discovering how to fine-tune NOMAD's tools to respond to industry needs was actually a main scope of the workshop.
- It is promoting a constructive resource- and information-exchanges between industry and research institutions, and to some extent among industry itself.
- It is encouraging and facilitating student placements as well as academic stuff embedding in industry (the workshop was, in the event, a venue where industry-generated job vacancies for young people with a background in data science could be announced and circulated).
- It is contributing to close the gap between academic and industrial research boosting communication and collaborations.

Programme

The workshop started with an introduction on the NOMAD project and a full update on its state of progress. The introduction outlined the objectives and activities of the NOMAD initiative as a whole; to make industry guests aware that analytics tools geared for practical data usage, are being produced and are currently already testable. (This is notably at variance with the US-based Materials Genome Initiative, where -so far- analytics does not play a noticeable role). We addressed the central theme of industry networking, with an overview of recent results achieved in an array of "Industry Case Studies" (carried out in direct collaboration with industry) and "Industry relevant use cases" (inspired by previously gathered industrial feedback).

The workshop turned to the key issue of usability of the data and graphics tools produced by the CoE, now testable with tutorial examples online.

A “Nomad’s Accomplishments so far” section of the workshop described the status of the different parts of the NOMAD project, and the challenges that still lie ahead. This namely covered (i) the NOMAD Encyclopedia which is the primary tool allowing access to the information stored within the database, already open and available to everyone as a fully open access tool. Two more short technical talks followed that l (ii) described the state of our “big-data analytics toolkit”, with practical examples of the functionalities that the CoE aims to provide to boost the exploitation of the materials database and (iii) describe the graphics tools planned in the CoE, based on remote visualization through a web-like interface. Virtual reality has also been discussed and exemplified as an aid tool for interacting with complex, high-dimensional data sets. A specific preliminary example of case study (iv) has been then showcased, involving a talk by a BP and a NOMAD young researcher (Ms Rachel Fort and Dr. Martina Stella, respectively) on the data aspects of tribochemical research on complex lubricants for automotive applications.

The above was followed by a third main session during which four key industry representatives laid out their specific needs and plans for computational materials science in general and their expectations for the NOMAD initiative in particular. This covered data-related R&D issue related to structural materials research, ranging from civil aviation to the transport, civil engineering, and energy sectors. A fourth session on the second day provided yet more practical examples of usage through the presentation of three more industry case studies ongoing in Spain, Denmark and Finland. We then moved on to gather feedback during two general panel discussions on
Industry Feedback and Suggestions (general) and Interaction With Industry (with recommendation gathered from the NOMAD Industry Advisory Committee). This discussion provided very useful feedback on how to shape all the forthcoming industry networking activities of this (and any similar) initiative, making it more relevant for industry users. The workshop was concluded with a public summary of the discussion and its outcomes.

Timetable

Day 1 – 05 Feb 2018

14:30 h Registration, coffee

Session I: Introduction

15:00 h Welcome and general introduction (Matthias Scheffler)

15:10 h NOMAD Industry Networking (Alessandro De Vita, Angel Rubio)

Session II: NOMAD's accomplishments so far

15:20 h The NOMAD Encyclopedia (George Huhs/Claudia Draxl)

15:30 h NOMAD data analytics (Luca Ghiringhelli)

16:10 h NOMAD advanced graphics (Rubén García Hernández)

16:20 h Coffee Break and Networking

Session III: Listening to Industry Needs
16:30 h David Rugg (Rolls Royce)
17:00 h Helmuth Sarmiento Klapper (Baker Hughes, TBC)
17:30 h Coffee Break
18:00 h Astrid Perlade (ARCELORMITTAL, TBC)
18:20 h David Gao (Nanolayers, TBC)
18:40 h Christoph Berndhäuser (Schott AG, TBC)
19:00 h Adjourn
19:30 h Conference Dinner

Day 2 –Tuesday, Feb. 06, 2018
08:45 h Coffee and networking
Session IV: Gathering Feedback
09:00 h Four Case Studies (Francesc Illas, Kristian Thygesen, Adam Foster, Martina Stella, Rachel Fort)
09.40 h General Discussion: Interaction with Industry
(on feedback gathering, software training, case studies, etc.)
10.00 h NOMAD-XT –Pillar 1: Claudia Draxl, "Exascale Computing and Software"
10.10 h NOMAD-XT –Pillar 2: Kristian Thygesen, "High-Throughput Screening and Workflows"
10.20 h NOMAD-XT –Pillar 3: James Kermode, "Big-Data Analytics and Data Processing"
10:30 h Coffee break

Session V: Summary
10.45 h Summary and Actions (Angel Rubio)
11:15 h Wrap Up and Farewell (Matthias Scheffler)
11.20 h Coffee, networking and end of NOMAD Industry Meeting

--- For the Industry Advisory Council and PIs only:
11:30 h IAC Meeting - PIs meet with members of the IACs for feedback/discussion
12.30 h Lunch and discussion

List of participants:

1. **Scheffler Matthias** Fritz Haber Institute of the Max Planck Society (FHI), Berlin
2. **Rubio Angel** Max Planck Institute for the Structure and Dynamics of Matter
3. **De Vita Alessandro** King's College London and University of Trieste
4. **Bromley Stefan** University of Barcelona (UB)
5. **Carbogno Christian** Fritz-Haber Institute (FHI)
6. **Compostella Michele** Max Planck Computing and Data Facility (MPG-MPCDF)
7. **Csanyi Gabor** University of Cambridge
8. **De Mier Torrecilla Monica** Barcelona Supercomputing Center (BSC)
9. **del Caño Teodosio** ONYX SOLAR ENERGY, S.L.
10. **Demuth Dirk** BASF SE
11. **Draxl Caludia** Humboldt-Universität zu Berlin (HUB)
12. **Echeverri-Restrepo Sebastian** SKF
13. **Fekete Adam** King's College London (KCL)
14. **Filip Sorin** BP Chemical Limited
15. **Fort Rachel** King's College London (KCL) - BP
16. **Foster Adam** AALTO University
17. **Gao David** Nanolayers
18. **Garcia Hernandez Ruben** Leibniz Supercomputing Centre
19. **Ghiringhelli Luca** Fritz-Haber Institute (MPG-FHI)

20. **Glawe Henning** Max Planck Institute for the Structure and Dynamics of Matter (MPG-MPSD)
21. **Goldsmith Bryan** Fritz-Haber Institute (FHI)
22. **Heinzel Stefan** Max Planck Computing and Data Facility (MPG-MPCDF)
23. **Helm Dirk** Fraunhofer
24. **Himanen Lauri** AALTO University
25. **Huhs Georg Barcelona Supercomputing Center (BSC)**
26. **Ignatius Janne** CSC – IT Center for Science Espoo
27. **Illas Francesc** University of Barcelona (UB)
28. **Kivioja Jani** Nokia
29. **Klinge Micheal** Springer Materials
30. **Koski Kimmo** CSC – IT Center for Science Espoo
31. **Kranzlmuller Dieter** Leibniz Supercomputing Centre
32. **Krein Micheal** Lockheed Martin
33. **Lampensherf Stefan** Siemens
34. **Lederer Hermann** Max Planck Computing and Data Facility (MPG-MPCDF)
35. **Levchenko Sergey** Fritz-Haber Institute (FHI)
36. **McKane Angela** BP Chemical Limited
37. **Milman Victor** Dassault Systemes BIOVIA
38. **Mohamed Fawzi** Fritz-Haber Institute (MPG-FHI)
39. **Moses Poul Georg** Haldor Topsoe A/S
40. **Nieminen Risto** AALTO University

42. **O'Brian Kylie** Pintail Services
41. **Pardini Lorenzo** Humboldt-Universität zu Berlin (HUB)
42. **Pavone Pasquale** Humboldt-Universität zu Berlin (HUB)
43. **Perlade Astrid** ARCELORMITTAL
44. **Poelking Carl** University of Cambridge
45. **Rampp Markus** Max Planck Computing and Data Facility (MPG-MPCDF)
46. **Riello Massimo** General Electric
47. **Rigamonti Santiago** Humboldt-Universität zu Berlin (HUB)
48. **Rinke Patrick** AALTO University
49. **Rubio Angel** Max Planck Institute for the Structure and Dynamics of Matter (MPG-MPSD)
50. **Rugg David** Rolls Royce
51. **Rupp Matthias** Fritz-Haber Institute (FHI)
52. **Sanyal Suchismita** Shell
53. **Sarmiento Klapper Helmuth** Baker Hughes
54. **Schaefer Angsar** BASF SE
55. **Scheffler Matthias** Fritz-Haber Institute (FHI)
56. **Shin Jungho** Humboldt-Universität zu Berlin (HUB)
57. **Sillanpää Atte** CSC – IT Center for Science Espoo
58. **Stella Martina** King's College London (KCL)
59. **Stokbro Kurt** Qunatum Wise
60. **Strange Mikkel** Technical University of Denmark (DTU)
61. **Sutton Christopher** Fritz-Haber Institute (FHI)

62. **Thygesen Kristian** Technical University of Denmark (DTU)

63. **Tomecka Daria** Fritz-Haber Institute (FHI

64. **Troppenz Maria** Humboldt-Universität zu Berlin (HUB)

67. **Valero Rosendo** University of Barcelona (UB)

65. **Vancea Ioan** Humboldt-Universität zu Berlin (HUB)

66. **Vatashavay Sriharsha** CSC – IT Center for Science Espoo

67. **Wimmer Erich** Materials Design

68. **Zastrow Thomas** Max Planck Computing and Data Facility (MPG-MPCDF)

69. **Ziletti Angelo** Fritz-Haber Institute (FHI)