

Dear Psi-k Community,

with this email we like to express our big THANKS for your support, and we like to ask for your continued assistance so that we can improve and advance NOMAD even further.

In November 2014, the PIs of the NOMAD (Novel Materials Discovery) Center of Excellence (CoE) launched the below noted survey. This stimulated quite some discussion, and initially some people were reluctant to the idea of sharing their input and output files. Meanwhile, the scientific culture has drastically changed, and the success of NOMAD was much bigger than expected: Initially, we expected to count maybe 100,000 calculations by the end of 2018. By now the NOMAD Repository and Archive already holds more than 50 million open access calculations!

One year ago, a *Nature Editorial* summarized the dilemma of *Open Science* (<https://www.nature.com/news/empty-rhetoric-over-data-sharing-slows-science-1.22133>): ***Empty rhetoric over data sharing slows science. - Governments, funders and scientific communities must move beyond lip-service and commit to data-sharing practices and platforms.***

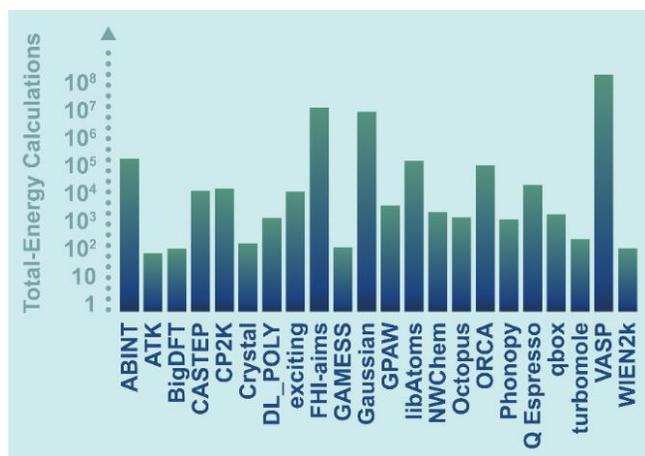
Indeed, for computational materials science NOMAD had changed this situation already.

The NOMAD CoE developed what is by now the largest repository for input and output files for the wider Psi-k community, and, since more recently, we also offer the services to force-field codes. NOMAD is now supporting 40 (!) different codes, and, in case that a code is not yet supported, we offer extensive help for writing the necessary parser. We just need to hear from you. A simple summary of the NOMAD Repository can be watched here: <https://youtu.be/UcnHGokl2Nc> -- see also <https://repository.nomad-coe.eu/>.

NOMAD serves the whole ecosystem of important computer codes of the Psi-k field.

- The uploaded data is checked for its consistency, and open access uploads (only those) will be processed and stored in the NOMAD Archive.
- The uploader can make the files open access immediately or share it with a few colleagues only. Open access can be delayed by up to 3 years.
- Open-access data are subject to the *Creative-Commons License*.
- With just a mouse click a *Digital Object Identifier (DOI)* can be requested. This makes the data citable.
- For downloading Open Access data no identification is necessary.

NOMAD is more than the largest repository! It helps you and your group, and it helps the whole community. It also developed parsers to transform the information provided into a code-independent and normalized format such that output from different codes can be compared. Obviously, only the open access data are being processed. This is the NOMAD Archive. The biggest contributors to the Repository & Archive are shown in the figure below:



Basically all important electronic structure codes added the below stamp to their code web pages:



The whole concept of the NOMAD CoE is described in this 3-minute movie:

<https://youtu.be/yawM2ThVIGw> and in the upcoming MRS Bulletin paper “NOMAD: The FAIR Concept for Big-Data-Driven Materials Science” (<https://arxiv.org/abs/1805.05039>).

Please, let us know your needs and provide suggestions how the services can be improved. Let us add that we are in the process of making the NOMAD Repository and Archive becoming a key element of the independent association and charity **FAIR Data Infrastructure**. More on this will be described later.

Thanks for your support over the last 3.5 years, and let us improve and advance NOMAD together even further.

With best wishes

Claudia and Matthias for the whole NOMAD Team (<https://nomad-coe.eu/the-team/team>)

On 11/21/2014 8:16 AM, Psi-k wrote:

Survey on Open Access of Data - all data

Dear CECAM and Psi_k community!

Please read this NOW and don't suspend it to tomorrow.

Our community has been carrying out CPU-intensive calculations for several years, but almost every group still stores their data on their own hard disks. Most calculated data are not even used and thrown away. In

several cases, high-throughput studies are performed, but typically, the data are just used for simple queries: Search if a certain property had been calculated already.

We must do more! If we could bring all the data from different groups together, we all would profit. For example, big-data analytics will become possible. Clearly, a first step would be to have access to suitable repositories, to enable

open access of data -- of all data (the complete input and output files).

We all will profit from this, and we kindly ask you to express your opinion on open access.

Why is open access (or sharing) of data appropriate, useful, and important?

Let us note this in bullet form:

- Open access of data implies that data can be used by anyone, not just the experts who develop or run advanced codes. If our data were openly available, many more people will work with the data, e.g. computer scientists, applied mathematicians, "analytic" condensed matter scientist, and more. We will be surprised what people will do with data, probably using tools that the present computational materials community does not even know.
- Many systems are being calculated again and again. If the full input and output files were available, much of repetition could be avoided.
- Typically, data will come together with a publication where the author had used them first. The availability of the data will be mentioned in the publication, and this publication will be cited by others who use the data. Thus, your citation index will improve.
- Nearly all computations in the field are supported by taxpayer's money. It should be a duty to publish the results - all the results.
- Rules of good scientific practice set by several science agencies, worldwide, already require to keep scientific data for 10 years. A good repository would practically help us and our students recall what was actually done some years ago.

In summary: We propose "a change in scientific culture" of computational materials science and engineering. It will help everyone, and it will significantly advance the field. Several repositories are considering the above noted issues, e.g. <https://cmr.fysik.dtu.dk/> or <http://nomad-repository.eu/>, just to name two examples.

If you support the open access idea, please click here:

Yes, I support this idea

If you consider open access of data and sharing a wrong move, please click here:

No, this is a wrong move

The following group of PIs asks you to support the above-sketched initiative because it will also enable a significantly advanced "big-data analytics" which will bring computational materials science and engineering forward by a significant step.

Matthias Scheffler (Fritz Haber Institute of the Max-Planck Society, Berlin, proposal coordinator), Claudia Draxl (Humboldt-Universität, Berlin), Daan Frenkel (University of Cambridge), Francesc Illas (University of Barcelona), Risto Nieminen (Aalto University, Helsinki), Angel Rubio (Max Planck Institute MPSD, Hamburg)

and UPV/EHU, San Sebastián), Kristian Thygesen (Technical University of Denmark, Lyngby), Alessandro De Vita (King's College London)

Please read a short summary of this initiative here: [here](#)

Thanks a lot, on behalf of all PIs!
Matthias