Manifesto of the NOMAD Implementation Network

Description of NOMAD

The Novel Materials Discovery (NOMAD) Laboratory maintains the largest worldwide repository for input and output files of all important computational materials science computer programs. Based on its open-access data, it builds several Big-Data Services helping to advance materials science and engineering by recognizing that data is a crucial raw material of the 21st century. The NOMAD Repository started in 2013 as a joint endeavour of the Humboldt-Universität Berlin and the Fritz-Haber-Institut der Max-Planck-Gesellschaft in Berlin, and then advanced into the NOMAD European Centre of Excellence (CoE), https://nomad-coe.eu, established in fall 2015 by bringing together eight research groups and four high-performance computing (HPC) centres. Now, NOMAD is making steps towards sustainability by starting a nonprofit association (a Berlin based "gemeinnütziger eingetragener Verein (e.V.)"). Pillar A of this e.V. covers 3 of the five NOMAD CoE components, i.e. the Repository (raw data), the Archive (normalized data), and the Encyclopedia (graphical user interface for the presentation of the data). This "NOMAD Pillar" aims to become a GO FAIR IN.

The original idea of the NOMAD Repository was to achieve an extensive sharing of scientific data. The concept was orthogonal to other data-base activities in several aspects: First, from the very beginning NOMAD did not just asked for (a subset) of results but for the full input and output files as well as detailed information about the computer program that was used. Second, NOMAD was never restricted to just a single simulation program or a restricted community but it intended to include all relevant programs, and it addresses all researchers of the whole field. This innovative approach for sharing represented a cultural shift in the scientific community, and it was a forerunner of what later became the FAIR data concepts. The mentioned initial concepts are still valid. NOMAD keeps data for at least 10 years (i.e. 10 years after the last upload are guaranteed), thus NOMAD is creating a stable and referenceable data domain for many decades. Data can become open access immediately after the upload or with a delay (embargo) time of a maximum of three years; during this period such data sets can be shared with selected people or referees of a journal. About 99% of all current data in NOMAD are in fact already open access. NOMAD issues DOIs, thus making data citable. Content uploaded to NOMAD is made available subject to the rules of the Creative Commons Attribution 3.0 License. People who upload data have to register, but for searching and download no registration or identification is necessary.

NOMAD not only collects computational material science data from many different labs worldwide, but also cleanses and normalizes it. That way mining this data becomes possible and opens perspectives towards finding structure, correlations, and novel information that could not be discovered from studying smaller data sets. Normalization not only means making the different data sets independent from the underlying materials-science computer program that created it, but to also making it stepwise FAIR compliant. These

measures, providing innovative tools and offering curated data enable researchers in basic science and engineering to identify new physical phenomena, and help industry to improve existing and develop novel products and technologies.

The simulation program-independent presentation of the NOMAD Archive relies on unified metadata, that were developed in close contact with the wider community, to label the data unambiguously. For further reading, see

L.M. Ghiringhelli, C. Carbogno, S. Levchenko, F. Mohamed, G. Huhs, M. Lueders, M. Oliveira, and M. Scheffler, Towards a Common Format for Computational Materials Science Data. Published as "Ψk Scientific Highlight of the Month, n. 131 (July 2016).
https://th.fhi-berlin.mpg.de/site/uploads/Publications/Psik_Highlight_131-2016.pdf ;

L.M. Ghiringhelli, C. Carbogno, S.V. Levchenko, F. Mohamed, G. Huhs, M. Lueders, M. Oliveira, and M. Scheffler,
Towards efficient data exchange and sharing for big-data driven materials science: metadata and data formats. npj Computational Materials 3, 46 (2017).
https://th.fhi-berlin.mpg.de/site/uploads/Publications/s41524-017-0048-5.pdf

In addition, NOMAD is working on an Encyclopedia that represents a user-friendly, public access point to the extensive knowledge contained in the NOMAD Archive. It will help interested users to use the NOMAD Archive. NOMAD is also conducting extensive outreach to industrial and academic end-users to achieve maximum impact and benefit, to engage even more labs worldwide in contributing and to change data practices in the labs.

Method of Work

The method of work is as follows:

- Aggregating relevant results in form of data (full input and output files of computational materials science studies) from different labs worldwide into the NOMAD Repository.
- Cleansing and normalizing these data sets, assigning PIDs in form of DOIs and unified metadata descriptions - all to be stored in the NOMAD Archive.
- Offering this archive to the interested materials-science community and beyond, based on free and open access.
- Offering help & support and guidance via an Encyclopedia, a user forum, courses and seminars to interested users with the intention to also improve data practices.

Reason to participate in GO FAIR

Since all NOMAD facilities are meant to be available for many years we want to adopt the state-of-the-art

methods in data organization and description. Therefore we want to make our systems FAIR compliant and not only apply current FAIR metrics but also contribute to their advancements.

To implement the FAIR principles we will adopt the Digital Object model pushed ahead by the C2CAMP Implementation Network, i.e. we see GO FAIR as an excellent basis to interact with other INs and to contribute actively to progress towards a more stable domain of digital data.

Guiding Purpose: Increased Interoperability through the stepwise implementation of the FAIR principles where FAIR DMP tools, Digital Objects and other GO FAIR results will play a role.

Deliverables

- Within GO FAIR NOMAD will present the state of the work and will apply FAIR metrics - once ready to us - to assess the FAIRness of its archive.
- NOMAD is willing to act as an early test candidate for FAIRmetrics to help optimizing the guidelines.
- NOMAD is eager to further develop its concepts together with other GO FAIR INs to achieve the coherence in data management and re-use.

Overarching Principle of Operation

We commit to comply with the Rules of Engagement of GO FAIR Implementation Networks.

The NOMAD core network is thus a network of experts that share the same objectives and it includes a larger number of experts as data scientists, developers, HPC experts etc. With respect to GO FAIR, NOMAD will be represented by Peter Wittenburg.

Members of the NOMAD leadership are

- Matthias Scheffler (Coordinator) FHI Max Planck Society Berlin
- Claudia Draxl (Deputy Coordinator) Humboldt University Berlin
- Peter Wittenburg (GOFAIR Liaison) MPCDF, Max Planck Society, Garching
- Raphael Ritz MPCDF, Max Planck Society, Garching
- Hermann Lederer MPCDF, Max Planck Society
- Markus Scheidgen FHI-MPG and HU Berlin

Further collaborators and advisors include

- Jose Maria Cela Barcelona Supercomputing Centre
- Nicolas Fabas MPCDF, Max Planck Society, Garching
- Luca Ghiringhelli FHI-MPG Berlin

- Xavier Gonze University Louvain
- Geoffroy Hautier University Louvain
- Janne Ignatius CSC Finland
- Francesc Illas University of Barcelona
- James Kermode Warwick University
- Kimmo Koski CSC Finland
- Dieter Kranzlmüller Leibniz Supercomputing Centre, Garching
- Markus Rampp MPCDF, Garching
- Gian-Marco Rignanese University Louvain
- Patrick Rinke Aalto University Finland
- Angel Rubio MPI for the Structure of Dynamics of Matter, Hamburg
- Jungho Shin FHI-MPG and HU Berlin
- Kristian Sommer Thygesen Technical University of Denmark, Lyngby
- Alessandro De Vita King's College London
- Thomas Zastrow MPCDF, Max Planck Society, Garching

The NOMAD leadership agrees with this document.

Signed by:

Matthias Scheffler Claudia Draxl Peter Wittenburg