NOMAD - a Research Data Management Platform for Materials Science

Andrea Albino

Physics Department and IRIS Adlershof, Humboldt-Universität zu Berlin, Berlin, Germany,

Data-driven material science has the potential to transform the way we design and develop materials. This emerging field represents a significant departure from traditional trial-and-error methods and empirical approaches that have characterized materials science for decades.

Finding predictive empirical relations that allow for precise control over various aspects of the experiments has posed a challenge to human cognitive abilities alone. This becomes even more complex when combining datasets from different labs or from different scientists due to the lack of established standards for data models and methods to capture the large number of experimental details, including elaborate workflows and a large diversity of instruments for characterization. Furthermore, linking data from different fields of materials science, such as synthesis, experimental characterization, and theory, using common metadata schemas and ontologies will foster new paradigms of discovery, engineering, and optimization of novel materials.

The FAIRmat (fairmat-nfdi.eu) project combines all the efforts in designing research data management (RDM) tools with the final aim of embracing the new data driven approach to the materials science [1].

Within FAIRmat, we develeop NOMAD (nomad-lab.eu) [2] as a platform and open-source software for making Materials Science data FAIR [3]. In its conception, NOMAD was originally designed as a data repository for publishing computational Materials Science data. The software is currently rapidly evolving to handle also experimental data and to play the role of a central research data management tool including a data analysis platform, too. Within this perspective, Electronic Laboratory Notebook (ELN) features and other tools are now available in NOMAD. The ELN is a digital tool used to replace traditional paper laboratory notebooks that allows users to design data schemas and to implement custom data visualization, combining manual inputs with automatic file parsing.

We started structuring data schemas for a particular set of experiments. Following a bottom-up approach, we are aiming to achieve a general description of the similarities recurring in each of them. These observations can lead to a common data structure that can be adopted as a standard for larger groups of users. Sharing a common data structure means being able to share or merge together information without any parsing barrier. Data modeling enables to meet the need of retrieving the available information for further analysis. Data retrieval can be performed in NOMAD with powerful search capabilities specifically designed for materials science in a graphical user interface (GUI) and with an API. Moreover, NOMAD can be installed in your local servers (aka NOMAD Oasis) guaranteeing your data privacy.

The application of the NOMAD platform on the whole RDM life cycle will be illustrated in this seminar, showcasing the available features and the underlying concepts and aims of RDM.

References

[1] Scheffler, M., et al. FAIR data enabling new horizons for materials research. Nature. 2022; 604, 635-642.

[2] Scheidgen et al. NOMAD: A distributed web-based platform for managing materials science research data. Journal of Open Source Software. 2023; 8(90), 5388.

[3] Wilkinson, M., et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016; 3, 160018.