# O 1: Tutorial: How to Use NOMAD's Workflow Utilities to Improve Data Management and Facilitate Discovery in Materials Science (joint session O/TUT)

NOMAD (nomad-lab.eu) [1] is an open-source, community-driven data infrastructure, that supports automated (meta)data extraction from a wide range of simulations, including ab initio and advanced many-body calculations as well as molecular dynamics simulations. NOMAD allows users to store both standardized and custom complex simulation workflows, which not only streamlines data provenance and analysis but also facilitates the curation of AI-ready datasets. This tutorial will focus on recently developed workflow functionalities and utilities within the NOMAD infrastructure. These advances enable high-throughput interfacing with the NOMAD repository, opening improved discovery pipelines by leveraging the benefits of NOMAD's comprehensive and FAIR-compliant data management system [2].

[1] Scheidgen, M. et al., JOSS 8, 5388 (2023).

[2] Scheffler, M. et al., Nature 604, 635-642 (2022).

Time: Sunday 16:00–18:00 Location: H3

**Tutorial** O 1.1 Sun 16:00 H3

**FAIR-data management with the NOMAD infrastructure: Core functionalities** — •Joseph F. Rudzinski — Physics Department and CSMB Adlershof, Humboldt-Universität zu Berlin, Germany

In this first part of the tutorial series, an overview of the NOMAD infrastructure will be provided. Attendees will learn how NOMAD processes raw data and stores it within a generalized data structure, and the corresponding GUI features that allow users to comfortably browse data. An example scenario will also be set up for use throughout the remainder of the tutorial series: *A researcher with a variety of data obtained within a project workflow would like to upload this data to NOMAD in order to link it to their manuscript while exposing the details of their (meta)data and retaining the scientifically relevant connections between the individual project tasks.*

**Tutorial** O 1.2 Sun 16:30 H3

**Using NOMAD's API for project management** — •Nathan Daelman — Physics Department and CSMB Adlershof, Humboldt-Universität zu Berlin, Germany

In this part of the tutorial series, you will learn how to interface with NOMAD programmatically using a Python module built to simplify the API (application programming interface). Functionalities for uploading data, editing metadata of uploads, creating datasets with multiple uploads, and publishing data will be covered. Attendees will use these functionalities to manage a portion of the data from the example project workflow, in particular, the subset of data that is automatically recognized and processed by one of NOMAD's existing parsers. *(For attendees without any Python experience, an alternative route to upload via the GUI will also be demonstrated!)*

**Tutorial** O 1.3 Sun 17:00 H3

**Creating custom entries in NOMAD using yaml schema and ELN integration** — •Andrea Albino — Physics Department and CSMB Adlershof, Humboldt-Universität zu Berlin, Germany

In this part of the tutorial series, attendees will learn how to create custom entries to store data that is not already supported by one of NOMAD's parsers. The basics of writing a schema, using NOMAD's ELN (electronic lab notebook) integration, and how to create simple plots of your data to visualize in the GUI will be covered. Attendees will then use this knowledge to manage the remainder of the data from the example project workflow, which is not automatically recognized by NOMAD.

**Tutorial** O 1.4 Sun 17:30 H3

**Creating custom workflow entries in NOMAD to link multiple uploads** — •Bernadette Mohr — Physics Department and CSMB Adlershof, Humboldt-Universität zu Berlin, Germany

In this last part of the tutorial series, attendees will complete the example project workflow storage by creating a custom workflow entry in NOMAD that connects all the uploaded tasks. The basics of the schema for defining custom workflows will be covered, followed by a demonstration of the straightforward creation of the required workflow file using the same workflow utility Python module as in the first part of the tutorial series. Finally, attendees will navigate NOMAD's interactive workflow graph visualizations to investigate the uploaded data, and learn how to obtain a DOI for their workflow.