Supplementary Information to:
Towards efficient data exchange and sharing for big-data driven materials science: Metadata and data formats

Luca M. Ghiringhelli[1,1], Christian Carbogno[1], Sergey Levchenko[1], Fawzi Mohamed[1], Georg Huhs[2,3], Martin Lüders[4], Micael Oliveira[5,6,2], and Matthias Scheffler[1,7]

[1]Fritz Haber Institute of the Max Planck Society, Berlin, Germany
[2]Barcelona Supercomputing Center, Spain [3]Humboldt-Universität zu Berlin
[4]Daresbury Laboratory, UK [5]University of Liège, Belgium
[6]Max Planck Institute for the Structure and Dynamics of Matter, Hamburg, Germany
[7]Department of Chemistry and Biochemistry and Materials Department, University of California–Santa Barbara, Santa Barbara, CA 93106-5050, USA

1 An example of accessing the Archive via NOMAD Resolve URI

The NOMAD Resolve URI is implemented as a REpresentational State Transfer (REST) API for the efficient browsing of the parsed and normalized data. A URL can be constructed to access all metadata in the archive via a web-browser or any other tool for sending HTTP requests. This is done by concatenating the string:
`https://analytics-toolkit.nomad-coe.eu/api/resolve/`
with a NOMAD Resolve URI. This URI is a path, consisting of the following parts, connected with slashes (`/`):

1. the unique identifier (UID) of one raw-data zip-archive,
2. the UID of one calculation contained therein,
3. a chain of `section-name/section-identifier/` - pairs, starting with `/section_run/`
4. the metadata name of a concrete value

For example, the following URL accesses the `energy_total` (metadata for the total energy for a given configuration of atoms and a given physical model) of one specific energy evaluation:
`https://analytics-toolkit.nomad-coe.eu/api/resolve/N-TULHlZnc9cnbg7ihzUALIlSdyww/C-DWGvyqvK2g_1yLyJf8nN3j_M-xd/section_run/0c/section_single_configuration_calculation/0c/energy_total/0c`
where the identifier "0c" means the first item in a list, e.g., the first `section_single_configuration_calculation` of possibly many such sections. The second item in the list would be identified by "1c", and so on.

In the list above, only the first part is mandatory. If the path points to a collection of data (an archive, a calculation, or a section), only structural information is returned. The complete content of the collection can be retrieved by appending the method `?format=recursiveJson`.

[1]ghiringhelli@fhi-berlin.mpg.de
[2]mjt.oliveira@ulg.ac.be

For instance, the URL:
`https://analytics-toolkit.nomad-coe.eu/api/resolve/N-TULHlZnc9cnbg7ihzUALIlSdyww/C-DWGvyqvK2g_1yLyJf8nN3j_M-xd/section_run/0c?format=recursiveJson`
displays the content of a specific calculation, which, e.g., is performed with the code "VASP" (value of the metadata `program_name`).
It is a "geometry optimization" (`section_sampling_method`), performed with the "PBE" exchange-correlation functional (`section_XC_functionals` in `section_method`), and it contains a sequence of configurations with an energy evaluation for each. At each level, one can read the field "uri" that, copy-pasted in the address line of the browser, would allow for direct access to that specific section content or concrete metadata value. In practice the unique identifiers of archives and calculations will be results of a queries, and automatic tools will deal with them, rather than users manually typing. Then the here presented API allows direct access to the information of interest for the full dataset retrieved by the query.

A full documentation of the REST API will be presented in a forthcoming publication.