Generating a FAIR crystal-structure database with the AiiDA informatics infrastructure

Marnik Bercx[1], Sebastiaan P. Huber[1], Nicolas Hörmann[1], Giovanni Pizzi[1], and Nicola Marzari[1]

[1]*Theory and Simulation of Materials (THEOS), and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland*

Computer simulations that use powerful electronic-structure techniques are nowadays widely used to characterize or predict materials' properties. Such efforts rely on databases of measured or calculated data, with structural data being especially useful. Here, we develop and validate a set of protocols to generate a comprehensive structural database of 3D materials abiding to the FAIR data principles. We start from structures taken from three major experimental databases: the Pauling file (MPDS), the inorganic crystal structure database (ICSD), and the crystallography open database (COD). After removal of non-stoichiometric compounds and duplicates, the 72,662 unique structures are refined with density-functional theory calculations using Quantum ESPRESSO, enabled by the open-source SIRIUS accelerated library.

Due to the large number of high-throughput calculations required to obtain the ground-state geometry of the structures, an allocation on a Tier-0 HPC facility was essential for this project. We successfully applied for a project in the 21st PRACE call (ELGEN, proposal 2020225458) on the GPU partition of CSCS's Piz Daint supercomputer, as SIRIUS has demonstrated an optimal performance on GPU architectures.

Since calculations are driven by the AiiDA (http://aiida.net) materials' informatics infrastructure, all curated workflows, the entire provenance of the simulations and the resulting structural data can be shared openly on the Materials Cloud (https://www.materialscloud.org/discover/mc3d). We present our protocols and their validation, together with the use of AiiDA's advanced automation and error handling features to create robust workflows for electronic-structure simulations. Moreover, all reproducible turn-key workflows can be run for an arbitrary structure using the AiiDAlab platform (https://www.materialscloud.org/work/aiidalab), an intuitive simulation environment that is easy to use for both computational and experimental scientists. The combination of automated, robust workflows designed by experts in the field with user-friendly interfaces that make these tools accessible to researchers in both academia and industry will be a powerful driving force for the discovery of new materials.