

Materials Cloud (201-003)

As an open, international, and fully FAIR repository of computational data and workflows

Materials Cloud (<http://www.materialscloud.org>) is a **web platform for Computational Open Science (see Fig. 1)**, conceived to a) assist Computational Scientists in the full life-cycle of their projects, starting from data generation, and b) track, visualize, and publish the controlled conditions under which the data was generated, i.e., record and share of the data provenance (the logical chain or, more precisely, graph of all input data and calculations that were used to produce the final result).

The core of Materials Cloud is a data generation platform, strongly coupled with an archive and a sharing platform for open research data, assisting scientists in data dissemination, curation and publishing with DOIs. This sharing platform, thanks to a rich web interface, allows users to browse the data, navigate its provenance, visualize it, and download it for further reuse, making data fully accessible and interoperable. In particular, the web-based data generation platform is named "**AiiDA lab**" (<https://www.materialscloud.org/aiidalab>) and is powered by our software framework AiiDA (<http://www.aiida.net>), that provides automation of simulations, of workflows and a completely automated provenance tracking to ensure reusability and reproducibility. AiiDA has a collection of plugins supporting over 100 simulation tools, and implementing over 80 advanced scientific workflows (see AiiDA plugin registry: <http://aiidateam.github.io/aiida-registry>). Thanks to the unique combination of features of AiiDA and Materials Cloud, our platform can ensure FAIR data sharing. Online since late 2017 and hosted at CSCS, Materials Cloud already makes hundreds of thousands of fully reproducible simulations openly available (<https://www.materialscloud.org/home#statistics>). To our knowledge, currently no other infrastructure exists that is designed to guarantee the same level of reproducibility and FAIR sharing of the results in the field of Computational Materials Science, Physics and Chemistry, coupled with automation of simulations and workflows.

The aim of this project is to scale up our Materials Cloud infrastructure, focusing in this 2021 extension mostly on two components: the AiiDA lab simulation platform, and the Materials Cloud Archive for long-term storage of research data. Our aim is to match the needs of the whole scientific community, in Switzerland and beyond, and specifically targeting simulation laboratories that work in close collaboration with experimental researchers. At the same time, we want to ensure that the whole infrastructure is scalable, secure, and easy to redeploy on the (bare-metal or virtual) hardware of any university or research lab.

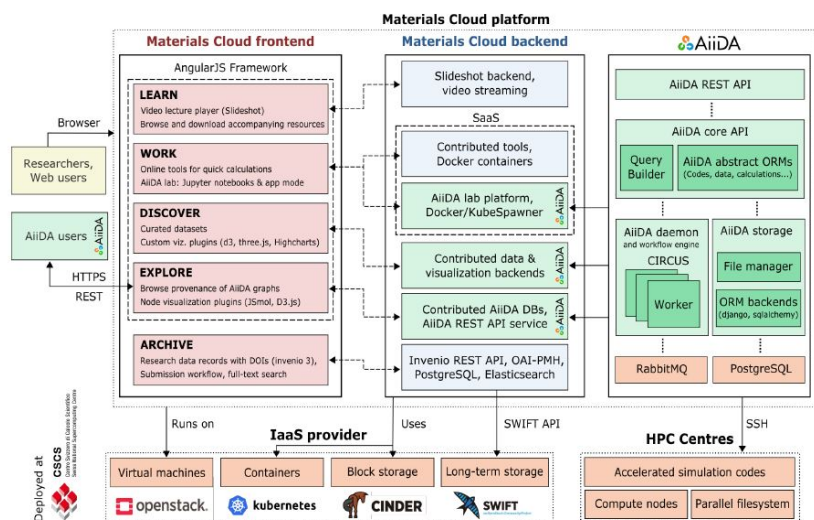


Fig. 1: The current Infrastructure of the Materials Cloud. A number of the sections in the backend (including AiiDA lab) are powered by AiiDA, that manages the execution of computational research on remote computers and HPC centres, and stores data and their provenance in a combination of databases and files. Its REST API, combined with the Materials Cloud REST API, provides the data displayed interactively by the AngularJS frontend. The long-term storage Archive portal relies on the robust Invenio framework (the same powering Zenodo at CERN) and stored data on the Swift object store at CSCS. Figure adapted from L. Talirz et al., arXiv:2003.12510 (2020).