

Materials Cloud (182-008)

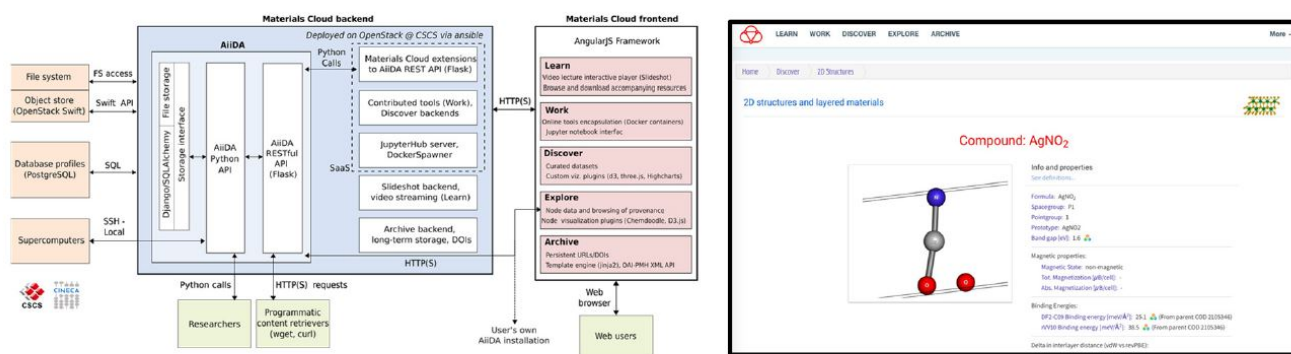
Materials Cloud as an open, international, and fully FAIR repository of computational data and workflows

Materials Cloud (<http://www.materialscloud.org>) is a web platform for Computational Open Science, conceived to assist Computational Scientists in the full life-cycle of their projects.

In general, we can distinguish three types of data: social, harvested and noisy; experimental, harvested in controlled conditions; and computational, generated in controlled conditions. Materials Cloud focuses on the last type. As such it focuses on two essential components: a) assist the researchers in the data generation part, 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).

Based on these principles, 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. The platform 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 of all computational research. Moreover, AiiDA has already a collection of plugins supporting over 20 simulation packages, and implementing various relevant scientific workflows. Thanks to the unique combination of features of AiiDA and Materials Cloud, our platform can ensure FAIR data sharing. Developed since 2015 and public since December 2017, it already provides data on materials computed by over 100'000 calculations. 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, coupled with automation of simulations and workflows.

The aim of this project is to scale up our Materials Cloud infrastructure to match the needs of the whole scientific community, in Switzerland first, extending then to Europe and beyond, while at the same time making sure that the infrastructure is stable, available at all times, flexible to support even more data types and scientific fields. In fact, while currently focused on Materials Science, the platform and the provenance model are very general and can be applied to many different disciplines, starting from computational physics and computational chemistry.



(Left) The current Infrastructure of the Materials Cloud. The backend is powered by AiiDA, that is able to connect to remote computational resources and stores data in a combination of databases and files (or, planned, to an object store). Its REST API, combined with the Materials Cloud REST API and some additional backend components, provides the data displayed interactively by the AngularJS frontend. (Right) A screenshot of one of the curated sections of Materials Cloud, with data always decorated with their full provenance, that can be interactively browsed ensuring reproducibility of research results.