AiiDA and the Materials Cloud: workflow engine with automated provenance and dissemination platform for Open Science

Sebastiaan P. Huber[1], Spyros Zoupanos[1], Leonid Kahle[1], Martin Uhrin[1], Nicolas Mounet[1], Rico Andreas Häuselmann[1], Snehal Kumbhar[1], Leopold Talirz[1,2], Aliaksandr Yakutovich[1,2], Elsa Passaro[1,2], Marco Borelli[1], Fernando Gargiulo[1], Ole Schütt[3], Andrea Cepellotti[1,4], Andrius Merkys[1,5], Boris Kozinsky[6], Berend Smit[2], Joost VandeVondele[7], Thomas Schulthess[7], Nicola Marzari[1], Giovanni Pizzi[1]

[1]Theory and Simulation of Materials (THEOS), EPFL, Lausanne, Switzerland
[2]Laboratory for Molecular Simulation (LSMO), EPFL, Sion, Switzerland
[3]nanotech@surfaces laboratory, Empa, Dubendorf, Switzerland
[4]UC Berkeley, USA
[5]Vilnius University, Lithuania
[6]Harvard University, USA
[7]Swiss National Supercomputing Centre (CSCS), Lugano, Switzerland

Abstract

Modern advances in computational technology have facilitated great strides in a wide variety of scientific disciplines and have led to the production of a wealth of valuable research data. However, the extraordinarily quick growth of computational capabilities has left the scientific world wanting for a simple yet effective way of managing these new workflows and the vast amount of data that they produce. We present AiiDA[1], a highly-automated and robust workflow engine written in Python, designed for high-throughput computational science. AiiDA automatically tracks data provenance (and stores it in the form of a directed graph) while managing and automating simulations running either locally or on supercomputers. All data and calculations are stored in a database and can be efficiently queried thanks to a simple but powerful query language. AiiDA enables computational science that is fully reproducible and facilitates sharing of research results. Its symbiotic counterpart, the Materials Cloud [2], is an interactive online platform powered by AiiDA. It is designed to enable seamless sharing and dissemination of resources, of raw data with their provenance as well as of curated open research data in computational science, and to provide cloud resources for simulations. The combination of AiiDA and Materials Cloud (whose architecture is shown in the figure) provides an Open Science
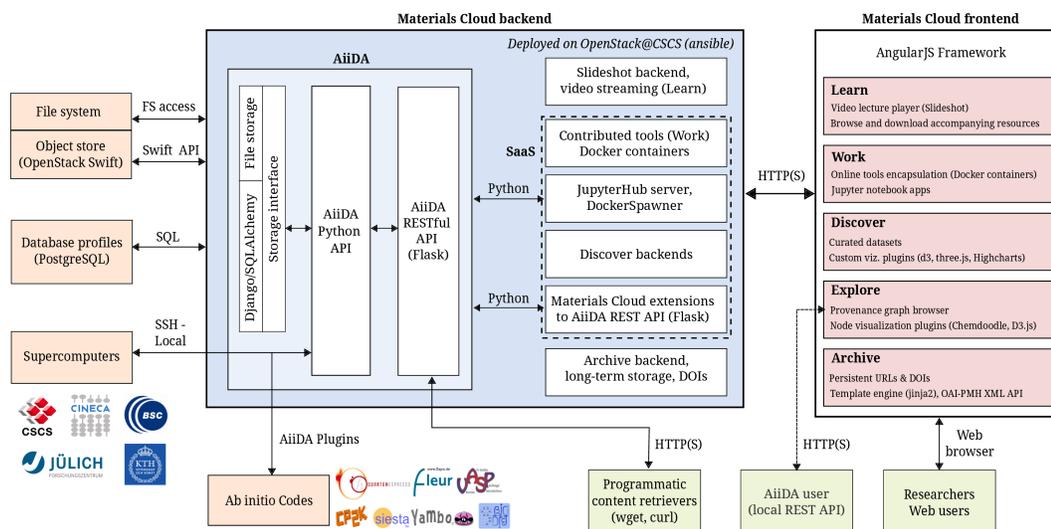
Figure 1: The architecture of AiiDA and the Materials Cloud. On the left, AiiDA can interact with remote supercomputers and stores its information both in a SQL database and as files. Moreover, it supports a number of external simulation codes via plugins. The AiiDA REST API, extended by Materials Cloud REST API, is one of the core components of the Materials Cloud backend, providing data to the frontend (that uses AngularJS).

Framework fully compliant with the FAIR (findable, accessible, interoperable and reusable) data principles [3]. The diligent and dedicated data management of AiiDA, combined with the user-friendly and ergonomic Materials Cloud, form an invaluable tool to any computational scientist.

References

[1] G. Pizzi et al., Comp. Mat. Sci. 111, 218 (2016) - www.aiida.net
[2] http://www.materialscloud.org
[3] M. D. Wilkinson et al, Sci. Data 3, 160018 (2016)