Manifesto of the Materials Cloud implementation network

Description of the Materials Cloud IN

The goal of the Materials Cloud IN is to provide the tools to help researchers across the world to ensure reproducibility of computational research and FAIR sharing of the resulting data, not only at the level of single calculations, but of the complete computational workflows that brought from the initial data to the final results. The Materials Cloud IN focuses currently on Computational Materials Science, but provides tools that can potentially be extended to other disciplines.

The Materials Cloud IN builds on top of a robust Open Science Platform for Computational Science, that rests on the following three pillars. 1) Widely-used, community-based open-source simulation engines, from local codes to massive HPC parallel codes running on supercomputers. 2) The open-source AiiDA materials informatics infrastructure, to manage, persist, share and reproduce complex computational workflows by automatically tracking the full provenance of data and calculations. 3) The Materials Cloud web platform, to enable sharing and dissemination of resources like: educational, research, and archiving tools; simulation software and services; curated and raw data.

These tools help accelerate research in the field of Materials Science, with direct societal impact on fields that are present in our everyday life, like: longer-lasting and safer batteries; more efficient energy harvesting materials (e.g., solar cells, thermoelectric materials) and catalysts; novel nanomaterials, functional materials and electronic devices.

The vision of the Materials Cloud IN nucleated in 2010, with the FET MARVEL Flagship Proposal, and took shape in 2013, with the Swiss MARVEL NCCR, and in 2015, with the H2020 MaX Centre of Excellence, that was later renewed in 2018 as the official European Centre of Excellence dedicated to Materials Science.

Currently, the Materials Cloud IN is supported by a consortium of partners and is open to further partnerships. These partners include the Swiss MARVEL NCCR, the European Centre of Excellence MaX, the swissuniversities P-5 Materials Cloud project, the European H2020 MarketPlace project, the European H2020 Intersect project the Swiss Platform for Advanced Scientific Computing PASC, and the EPFL Open Science OSSCAR project, involving currently a total of over 200 researchers. An updated list of partners can be found at on the Materials Cloud Partners page.

In addition, we are among the founding partners of the OPTiMaDe consortium, aiming at making materials databases interoperational by developing common formats and a common REST API specification.

Our Materials Cloud IN represents an unique network of partners in the field of Materials Science. The distinguishing feature of our network is that we are not only interested in collecting data and making it accessible but also in providing the core tools needed to make computational research open, reproducible and FAIR: 1) **open-source automated**

provenance-tracking workflow engines; 2) **robust and open turn-key workflows** for the computation of relevant materials properties; 3) **FAIR-sharing platforms** that not only make data Findable with DOIs, but provide Accessible web interfaces to them, ensure data is stored in Interoperable formats that are accepted by many different simulation codes, and help guarantee that shared data is not only Reusable but also Reproducible.

Our tools are backed by a lively and diverse "critical mass" community of developers of plugins for most of the codes of our scientific field, as demonstrated by the list available on the AiiDA plugin registry. Plugins are available for over 45 software packages and 90 code executables (as of March 2020). These are developed not only within the framework of the supporting projects, but also by independent researchers that were captivated by the core ideas and concepts of the Materials Cloud IN.

Overarching Objectives

- Developing ontologies to represent computational data and their provenance, and tools to **automate simulations** (both on local resources and on remote HPC clusters and supercomputers) and automatically **track the provenance of these data guaranteeing reproducibility**. We will leverage the existing strong networks like OPTiMaDe that is a neutral and lively platform, already including most of the stakeholders in the field.
- Assessing and possible extension of the Materials Cloud IN to other networks in the GO FAIR Community.
- **Developing robust FAIR workflows** for reliable and automated computation of materials properties using the most accurate and best-performant codes available nowadays.
- Assessing the current level of FAIRness of the platform, reducing from few to zero the places where human interaction is needed to access and understand the data.
- FAIRification and maintenance of the Materials Cloud platform as a place for researchers to **share their raw simulation data together with the full provenance** (EXPLORE section)
- Offering **curated data services**, where relevant materials properties, datasets and their **interlinked provenance** are made FAIR.
- Offer **archiving services** for all researchers in the field, where PIDs are assigned to each entry and data have guaranteed preservation and are made available for at least 10 years from their deposition. For data generated with AiiDA, corresponding Accessible EXPLORE sections will be interlinked to browse interactively the provenance of the simulations.
- Offer **cloud simulation services** using the AiiDA lab, making simple web GUIs available to everybody (not only experienced computational scientists), and develop the FAIR capability of existing tools for automated provenance tracking of the robust workflows developed, and facilitating the FAIR access of the results in the Materials Cloud.
- Work with the GO TRAIN Pillar to organize periodic **training** events to increase awareness of the FAIR principles in the scientific field and provide direct guidance on the use of the FAIR and open tools available. These are backed by a training platform on the LEARN section of Materials Cloud, where we disseminate openly videos of lectures and courses, and tutorials material including instructions together with virtual machines with all software already installed.

Reason to participate in GO FAIR

We believe that researchers in the field are aware of the principles of FAIR data and Open Science, but the effort to make data FAIR is typically too large without the suitable tools to assist. We are committed therefore to provide such tools to researchers and help them in adopting best practices and common formats.

We believe that the effort to make data FAIR cannot be applied "a posteriori" after the simulation have been run and published. Instead, the tools should assist researchers in proper data stewardship that begins before data is even collected, and continues while they are performing their simulations, keeping track in a fully automated way of the computational workflows and the full provenance of the data generated. Thanks to FAIR automatic provenance tracking, much higher guarantees of reproducibility can be offered with respect to recreating the provenance information at the end. With this approach, data access according to the FAIR principles becomes straightforward.
We have been committed since 2014 to provide open-source, robust and reliable tools and platforms available to the whole community. We want to participate as an implementation network of GO FAIR to contribute with our use cases and experience in the field, interact with other INs, increase interoperability of the tools across disciplines, and advance the domain of interlinked digital data with appropriate field-specific metadata, developed using common formats.

Specific Tasks

- Within GO FAIR, the Materials Cloud IN will represent the use cases of Computational Materials Science and automated workflow management and execution, helping to shape the FAIR Maturity Indicators to represent the FAIRness level of computational workflows and of computed data decorated with their provenance.
- Within the GO BUILD pillar, the Materials Cloud IN will continue to assess and develop the core tools needed for open and FAIR computational materials science, as well as developing a sustainable program to ensure the openness of the online portal and tools used to provide machine-accessibility of complex simulation data in the long term.
- Within the GO TRAIN pillar, the Materials Cloud IN will ensure the organization of periodic training events for computational materials scientists on writing reproducible workflows, on the tracking of data provenance in materials science, on FAIR data access and reuse. We will address not only in Europe but also a diverse set of countries (as already initiated in the past, see our Event page on the AiiDA website, with an average of 4 training events organized per year across the world).

Overarching Principle of Operation

We commit to comply with the Rules of Engagement of GO FAIR Implementation Networks.

The Materials Cloud IN core network is a network of experts in the field of Computational Materials Science, sharing the objective of making research fully reproducible and encouraging sharing of the results, with their complete provenance, according to the FAIR principles. We commit to make this possible via 1) the development of open tools critical to enable FAIR access of data and 2) the organization of training events across the world, also

outside of Europe, to develop the culture of the FAIR principles and of reproducible computational science.

Partners

- Stefano Baroni, SISSA (Italy)
- Stefan Blügel, Forschungszentrum Jülich (Germany)
- Arrigo Calzolari, CNR NANO (Italy)
- Michele Ceriotti, EPFL (Switzerland)
- Anne-Clemence Corminboeuf, EPFL (Switzerland)
- Thierry Deutsch, CEA (France)
- Andrea Ferretti, CNR NANO (Italy)
- Espen Flage-Larsen, SINTEF (Norway)
- Alberto Garcia, ICMAB (Spain)
- Luigi Genovese, CEA (France)
- Adham Hashibon, Fraunhofer (Germany)
- Sebastiaan P. Huber, EPFL (Switzerland)
- Andrea Marini, CNR ISM (Italy)
- Nicola Marzari, EPFL (Switzerland)
- Elisa Molinari, UNIMORE (Italy)
- Pablo Ordejón, ICN2 (Spain)
- Daniele Passerone, Empa (Switzerland)
- Carlo Pignedoli, Empa (Switzerland)
- Giovanni Pizzi, EPFL (Switzerland)
- Thomas Schulthess, ETHZ and CSCS (Switzerland)
- Berend Smit, EPFL (Switzerland)
- Atsushi Togo, NIMS (Japan)
- Joost VandeVondele, ETHZ and CSCS (Switzerland)
- Daniele Varsano, CNR NANO (Italy)
- Daniel Wortmann, Forschungszentrum Jülich (Germany)
- Oleg Yazyev, EPFL (Switzerland)