Demo of Materials Cloud and AiiDAlab: integrated platforms for FAIR computational materials science

Giovanni Pizzi (EPFL); Carlo Pignedoli (Empa)
DFT materials simulations

Our research: Materials simulations

Nature (2014): 12 papers on Density Functional Theory among the top-100 most cited papers in the entire scientific literature

Accuracy and predictive power of quantum engines

150,000x increase in the past 20 years

1 month (2001) → 10 seconds (2021)

DFT materials simulations

Our research: Materials simulations
Nature (2014): 12 papers on Density Functional Theory among the top-100 most cited papers in the entire scientific literature

Combined: materials design and discovery via high-throughput computations

Accuracy and predictive power of quantum engines

150,000x increase in the past 20 years

1 month (2001) → 10 seconds (2021)

Leverage supercomputers to compute and predict materials properties

Aim: Compute properties for all of them (and even new, invented ones) and discover novel functional materials
Challenges in *high-throughput HPC*

- **Workflow automation**
  - Need tools to define complex workflows with advanced error handling
  - An automated, robust and scalable engine to run the workflows

- **Data management**
  - Data should be stored reliably and efficiently
  - Stored data should be interoperable and queryable

- **Reproducibility**
  - All produced data should be reproducible by storing the full provenance
Aiida

- COMPUTATIONAL SCIENCE INFRASTRUCTURE
- FOR HIGH THROUGHPUT WORKFLOWS
- WITH FULL DATA PROVENANCE

Language: implemented and API in python
License: MIT open source http://www.aiida.net/
Source: https://github.com/aiidateam/aiida-core

- Scalable workflow engine
- Automated full data provenance
- Built-in support for HPC
- Flexible plugin system

G. Pizzi et al., Comp. Mat. Sci. 111, 218-230 (2016)
S.P. Huber et al., Scientific Data 7, 300 (2020)
Data provenance

Simple recipe

- Store data transformations or 'calculations'
- Store its inputs and their metadata
- Store its outputs and their metadata
- Most crucially store the inter-connections
Data provenance

Simple recipe

- Store data transformations or ‘calculations’
- Store its inputs and their metadata
- Store its outputs and their metadata
- Most crucially store the inter-connections

Provenance graphs

- When data gets reused, a directed graph is created
- That quickly grow in complexity even for “simple” workflows
Data provenance

Simple recipe

- Store data transformations or ‘calculations’
- Store its **inputs** and their metadata
- Store its **outputs** and their metadata
- Most **crucially** store the **inter-connections**

Graph requirements

- Needs to be automated
- Needs to be stored *as data is created*

**Complexity grows quickly even for simple workflows: impossible to reconstruct a posteriori** without a tool like AiiDA

*Graphical representation of actual AiiDA database*
AiiDA architecture
AiiDA tutorials

https://aiida-tutorials.readthedocs.io/
FAIR **data** access and sharing
FAIR sharing in AiiDA: data and graphs

- Share directly between private AiiDA instances
- Share on online repositories
L. Talirz et al., Scientific Data 7, 299 (2020)
Materials Cloud

- **AiiDA** is the ‘engine’, like **git** - used in production since 2015
- **Materials Cloud** is the dissemination platform (like **GitHub, GitLab, ...**) and more (cloud computing and data generation platform) - online since 2017
FAIR data sharing: Archive, Discover, Explore

Recommended data repository by Nature’s journal Scientific Data

Indexed by Google Dataset Search and by EUDAT/EOSC’s B2FIND

Registered on FAIRsharing.org and re3data.org

https://open-research-europe.ec.europa.eu/for-authors/data-guidelines
Interface for users submitting data

Submission form

User view (own entries)
Interface for users submitting data

Submission form

User view (own entries)

Ideal solution for data management
Open (and free) for any researcher in computational materials science
(5GB limit, 50GB when using AiiDA)
Open data sharing: Archive, Discover, Explore

Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds

Nicolas Mouren, Marco Gobbi, Philippe Schwalbe, Davide Campi, Andrea Merkys, Antuno Marrazzo, Thibault Söhne, Ivan V. Castelli, Andrea Cepellotti, Giovanni Pizzi, Nicola Marzari

1 Theory and Simulation of Materials (TUM), and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland
2 Virtus, University Institute of Biotechnology, Saulės al. 7, LT-10247 Vilnius, Lithuania

* Corresponding author email: nicolas.mouren@epfl.ch, nicola.marzari@epfl.ch

DOI assigned

Direct links to Discover & Explore

Data (and metadata) guaranteed to be online for at least 10 years after deposition
Open data sharing: Archive, Discover, Explore

**DISCOVER**

**Compound: MgBr₂**

- **Info and properties**:
  - Formula: MgBr₂
  - Spacegroup: P-3m1
  - Pointgroup: -3m
  - Prototype: CdI₂
  - Band gap (eV): 4.8

- **Magnetic properties**:
  - Magnetic State: non-magnetic
  - Tot. Magnetization [μB/cell]: -
  - Abs. Magnetization [μB/cell]: -

- **Binding Energies**:
  - DF2-C09 Binding energy (meV/Å²)
    - (From parent COD 9009107)
  - rWV10 Binding energy (meV/Å²)
    - (From parent COD 9009107)

- Delta in interlayer distance (vdW vs revPBE):
  - $\Delta_{vdW}$ [Å]: 17.1
    - (From parent COD 9009107)
  - $\Delta_{WV10}$ [Å]: 18.3
    - (From parent COD 9009107)

**UUID links to jump to the provenance graph in the EXPLORE section**

**EXPLORE**

- **Browse the full AiiDA provenance graph (inputs, outputs, …) at any level**
Data management plans (DMP) and FAIR principles

- Combination of **AiiDA + Materials Cloud** (Discover, Explore, Archive): **FAIR-compliant sharing**
  - **Findable**: DOIs with standardized metadata
  - **Accessible**: web interface to browse data, calculations and provenance, curated data in Discover section
  - **Interoperable**: data linked via the AiiDA directed graph; data structures reusable between different codes
  - **Reusable**: downloadable data, encourage open (CC) licences

  + **reproducible** thanks to full provenance

- We provide **DMP templates** for researchers using Materials Cloud

Below, we provide templates for data management plans using the Materials Cloud Archive (with and without AiiDA).

<table>
<thead>
<tr>
<th>Funding Body</th>
<th>DMP template (using AiiDA)</th>
<th>DMP template (no AiiDA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNF</td>
<td>.docx .odt .pdf</td>
<td>.docx .odt .pdf</td>
</tr>
</tbody>
</table>
Beyond FAIR *data* sharing: FAIR *simulation* access
Workflows to generate data

- Given a material, we compute advanced quantities: often non-trivial, result of a complex workflow

- AiiDA provenance graph: explain know how the structure was computed and to **reproduce that single specific calculation**: log of “what happened in the past”

- AiiDA workflow engine: Flexible python interface to encode complex scientific steps
  - Provides robustness (connection drops, ...) and allows to easily implement "error handlers" for common code-specific errors (crashes, non-convergence, end-of-walltime, ...)
  - Added bonus: automated provenance tracking and reproducibility
Mini-example of simple calculation submission

```python
PwBandsWorkChain = WorkflowFactory('quantumespresso.pw.bands')
code = load_code('pw.x@daint')
structure = StructureData(ase=ase.io.read('BN.xsf'))
builder = PwBandsWorkChain.get_builder_from_protocol(code=code, structure=structure)
... # Possible overrides
workchain_node = submit(builder)
```
FAIR sharing in AiiDA beyond data: codes, plugins and workflows

Plugins collected in the AiiDA plugin registry
• **100+ codes currently supported**, 120+ workflows
• Many are **community-contributed**

https://aiidateam.github.io/aiida-registry/
The need for turn-key solutions

- Like in a car: many end-users, drive without need to know how the engine works
  - Engines are "robust"
  - Just turn the key and drive

- I still need a driving license, but:
  - many needs addressed by driving my own car
  - no need to learn again to drive if I change car brand
A first goal

- As a non-expert, be able to ask

"Please run an equation of state with code [Quantum ESPRESSO|SIESTA|FLEUR|VASP|CASTEP|...] on the XXX supercomputer, using YY nodes, and automatically choose numerical parameters (basis set size/cutoff, k-points, ...) to get converged results."

- As an expert:
  - check the automatic parameters
  - adapt them if needed
  - check details of already-run simulations (by someone else): via provenance tracking
What are the challenges?

- **Have turn-key solutions** to calculate common materials properties
  Robust, with automatic error recovery, and automatic parameter choice

- **Accessible to non-experts, without losing control for experts**
  Automatic choice of numerical parameters

- **Interoperability** among several quantum engines
  Same input and output formats, and data types

Biggest challenge: all of the above at the same time
Code interoperability: common workflow interfaces

• Robust, cross-verified and accessible simulations with unified interface

https://github.com/aiidateam/aiida-common-workflows

• Effort started in the MaX CoE, but open to many more contributions
Code interoperability: common workflow interfaces

https://github.com/aiidateam/aiida-common-workflows/

$ aiida-common-workflows launch eos siesta --structure=Al --protocol=precise

Implementations for 11 quantum codes

EOS only for the 9 with PBC; relax and dissociation for all 11
From common workflow interfaces to common workflows

- We now have code-specific common workflow interfaces (e.g. for relaxation) with common interface
- We can now **implement (only once!) common workflows** using the workflows with common interfaces:
  - Dissociation curves
  - Equation of states (EOS)
  - Phonons
  - ...
Equation of states: results

- Tested (simple cases) for:
  - insulators
  - metals
  - magnetic systems
    - ferromagnetic
    - antiferromagnetic

Figure 7. EOS for Si, Al and GeTe. Results obtained with the code-agnostic EquationOfStateWorkflow. For each code, the energy is shifted to set the minimum energy to zero. The EOS has been computed with all codes discussed in this work, except ORCA and Gaussian, which are mainly specialized for non-periodic systems. In addition, for GeTe, results are missing for BigDFT, CP2K, FLEUR and NWChem (see Table II in the Supplementary Information for more details). The label QE stands for Quantum ESPRESSO.
Dissociation curve: results

H₂ dissociation curve

Distance between H atoms (Å)

Energy (eV)

-32.0
-31.5
-31.0
-30.5
-30.0
-29.5
-29.0
-28.5
-28.0
-27.5
-27.0

ABINIT
CASTEP
CP2K
Gaussian
ORCA
QE
SIESTA
VASP
Making it easy to use: Quantum Mobile

- It contains:
  - AiiDA
  - Quantum codes (all those that are free - some only work in serial for now)
  - AiiDA plugins and workflows (all, open-sources)
  - AiiDA common workflows code (since v21.05.1)
  - ...and much more
"Quantum Mobile" virtual machine (VM)

Testimonials

Compared to [courses of] previous years, the overhead due to technical problems and misunderstandings has been greatly reduced. Quantum Mobile is absolutely the right tool to have.

Stefaan Cottenier, compmatphys.org

Quantum Mobile is really a fantastic tool! I do think that it really gives a gigantic help for the organisation of Schools focusing on simulation codes.

Andrea Marini, founder of the Yambo code
Running in Quantum Mobile
Running in Quantum Mobile
Running in Quantum Mobile
All data is available and reproducible!

All data is available and reproducible!

You can import the data in the Quantum Mobile, and visualise it e.g. within Materials Cloud.


License

Files and data are licensed under the terms of the following license: Creative Commons Attribution 4.0 International. Metadata, except for email addresses, are licensed under the Creative Commons Attribution Share-Alike 4.0 International license.
Current plans: second phase of collaboration

Project now continuing with an even larger group of collaborators

- New codes (JuKKR, GPAW, Wien2K)
- Common workflow interfaces for band-structure and (P)DOS
- Common workflow interfaces for Born-effective charges and dielectric constant
- Common phonon workflow based on phonopy
- Cross-validation on 500+ oxides
- AiIDAlab GUI

Follow-up of the work of S. Cottenier (see talk of yesterday), comparing of various oxide configurations across all codes involved

Not only testing accuracy, but also releasing automated recipes to run with that accuracy!
AiiDAlab

- Comes with a preconfigured AiiDA setup, **ideal interface for turn-key workflows** to run directly in the browser
- Custom **AppMode** extension to make notebooks look & feel like real web apps: **only python for development, no coding skills to use**
- Easy to redeploy: JupyterHub + DockerSpawner on bare metal or VMs, but also kubernetes for scalability
Easy access to these advanced HPC capabilities to everybody

https://www.materialscloud.org/work/aiidalab
Making simulations accessible: AiiDAlab experience at EMPA (CH)

https://www.youtube.com/watch?v=x4mWFjSoAO4
https://www.materialscloud.org/work/aiidalab
AiiDA and Materials Cloud teams

Contributors for the 40+ plugins: Quantum ESPRESSO, Wannier90, CP2K, FLEUR, YAMBO, SIESTA, VASP, CASTEP, CRYSTAL, …

Contributors to aida-core and former AiiDA team members —
Oscar Arbelaez, Michael Atambo, Valentin Bersier, Marco Borelli, Jocelyn Boulleïer, Jens Bröder, Ivanov E. Castelli, Andrea Cepellotti, Keija Cui, Vladimir Dikan, Marco Dorigo, Y.-W. Fang, Fernando Gargiulo, Marco Gibertini, Davide Grassano, Dominik Gresch, Conrad Johnston, Rico Häuselmann, Daniel Hollas, Eric Hontz, Jianxing Huang, Christoph Koch, Espen Flage-Larsen, Ian Lee, Daniel Marchand, Antimo Marrazzo, Andrius Merkys, Simon Pintarelli, Nicolas Mounet, Tiziano Müller, Gianluca Prandini, Philip Rüßmann, Riccardo Sabatini, Ole Schütz, Philipppe Schwaller, Andreas Staminger, Atsushi Togo, Daniele Tomerini, Nicola Varini, Martin Uhrin, Jason Yu, Austin Vadoks, Bonan Zhu, Mario Zic, Spyros Zoupanos

Common workflows for computing material properties using different quantum engines

Sebastiaan P. Huber1,2, Emanuele Bosoni3,4, Marnik Berce5,6,7, Jens Bröder8,9, Augustin Degomme10, Vladimir Dikan11, Kristjan Eimre12, Espen Flage-Larsen13,14, Alberto Garcia15, Luigi Genovese16,17, Dominik Gresch18, Conrad Johnston19,20, Guido Petretto21, Samuel Poncé1, Gian-Marco Rignanese22, Christopher J. Sewell1, Berend Smit13, Vasily Tsiplyavtsev14,23, Martin Uhrin24, Daniel Wortmann25, Aliaksandr V. Yakutovich26,27, Austin Zadoks1, Pezhman Zarabadi-Poor28,29, Boan Zhu30,31,32, Nicola Marzari33, and Giovanni Pizzi34

+all new collaborators for the follow-up (currently running) project!
Acknowledgements and funding

**SNSF NCCR “MARVEL”**
Discovery of new materials via simulations and dissemination of curated data

**H2020 Centre of Excellence “MaX”**
Scaling towards exascale machines and high-throughput efficiency

**Swissuniversities P-5 “Materials Cloud”**
Scaling the web platform, extending to more disciplines

Moreover:

- **H2020 NEP** (starting 2021)
- **nffaeu**
- **EMMC**
- **BIG-MAP**
- **INTERSECT**
- **H2020 OpenModel** (starting 2021)
Summary

- AiiDA+Materials Cloud: FAIR access to simulation data
- In addition: AiiDA+common workflows: FAIR simulations
  - Open-source, robust turn-key workflows for materials properties
  - Common interface among 11 quantum engines (more coming)
  - Easy to use, but flexible for experts, designed to be interfaced to GUIs
  - Basis of verification studies (oxides set), and starting point for more common workflows