Center for Scientific Computing, Theory and Data

Beyond FAIR data: FAIR and reproducible workflows with AiiDA, Materials Cloud and AiiDAlab

Giovanni Pizzi, PSI

EnhanceR symposium, 7 November 2024

My group's role ("Materials Software and Data")

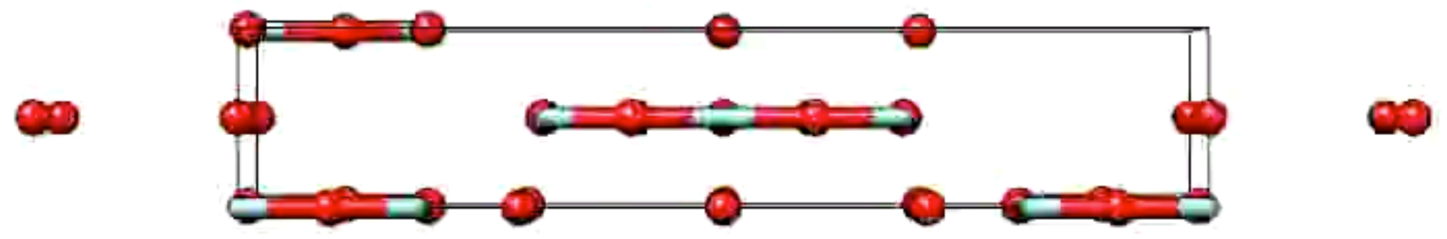
PSI

Between research and research software engineering (RSE)

- 11 people in my group, **~half are RSE**
- Myself: physicist. Most people in the group (both students/postdocs and RSE): **physics/chemistry/mat. science background**

- **Goal**: develop tools that enable:
  - **efficient** high-throughput research
  - automated simulations with **reproducibility**
  - **robust automated workflows** with minimal input/knowledge required
  - **seamless access** to advanced workflows by "non-experts" (e.g. via GUIs)

Scientific goal: Compute materials properties with supercomputers



Aim: Compute properties for all of them
(and even new, invented ones)
and **discover novel functional materials**

Challenges in high-throughput HPC

PSI

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

Further challenges to make it FAIR

PSI

FAIR data

- Make it easy for users to generate (and publish) FAIR data with **minimal effort for researchers**
- Can we leverage ontologies/semantic annotations to make data **machine-interoperable**?

Beyond FAIR data: FAIR workflows

- Can we make workflows **interoperable?** (e.g. between codes)
- Can we make workflows and advanced simulation methods **accessible?**

Reproducible simulations and interoperable workflows

AiiDA

PSI

- **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

MIT LICENSED

open source initiative®

Scalable workflow engine: ***robustness***

Automated full data provenance: ***reproducibility***

Built-in support for HPC: ***performance***

Flexible plugin system: ***interoperability***

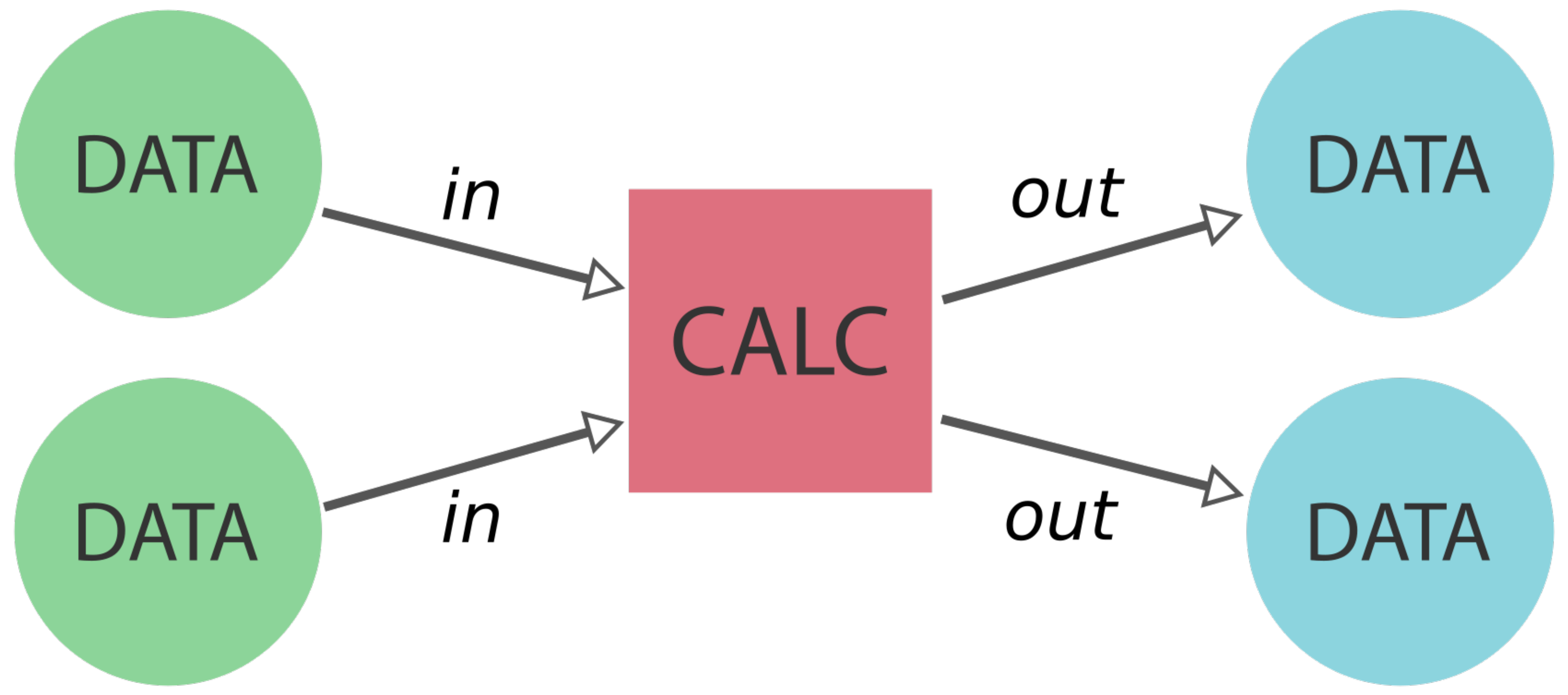
G. Pizzi et al., Comp. Mat. Sci. 111, 218-230 (2016)

S.P. Huber et al., Scientific Data 7, 300 (2020)

PSI

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

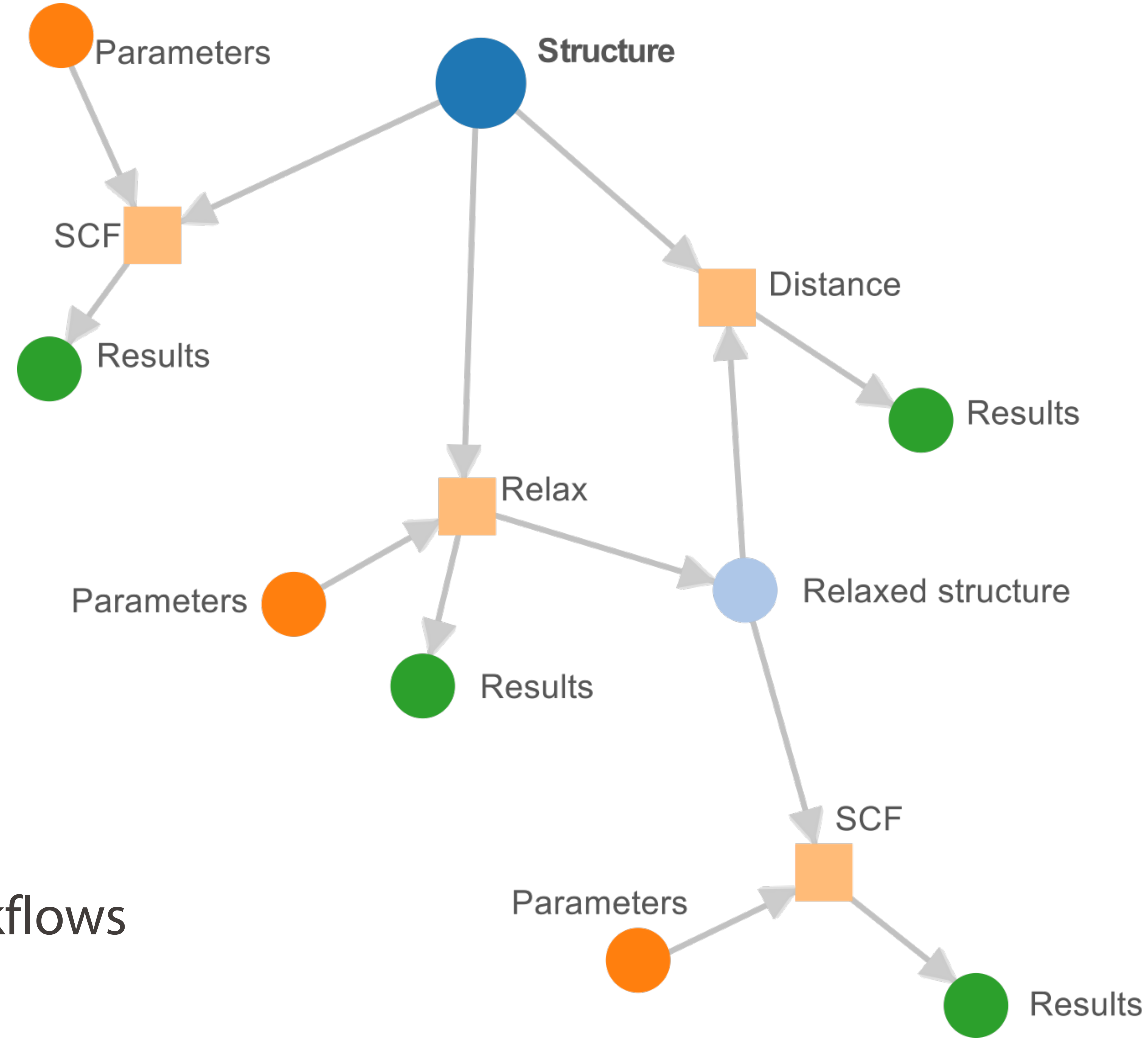
PSI

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

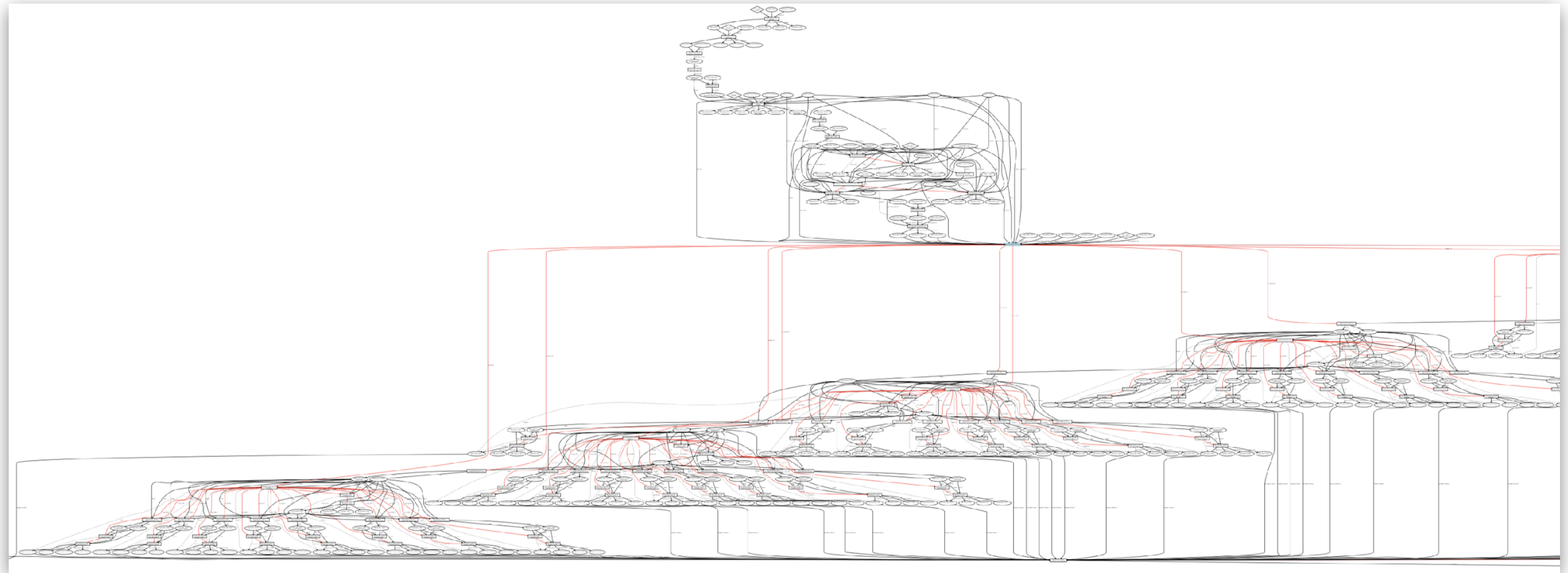
Simple recipe

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

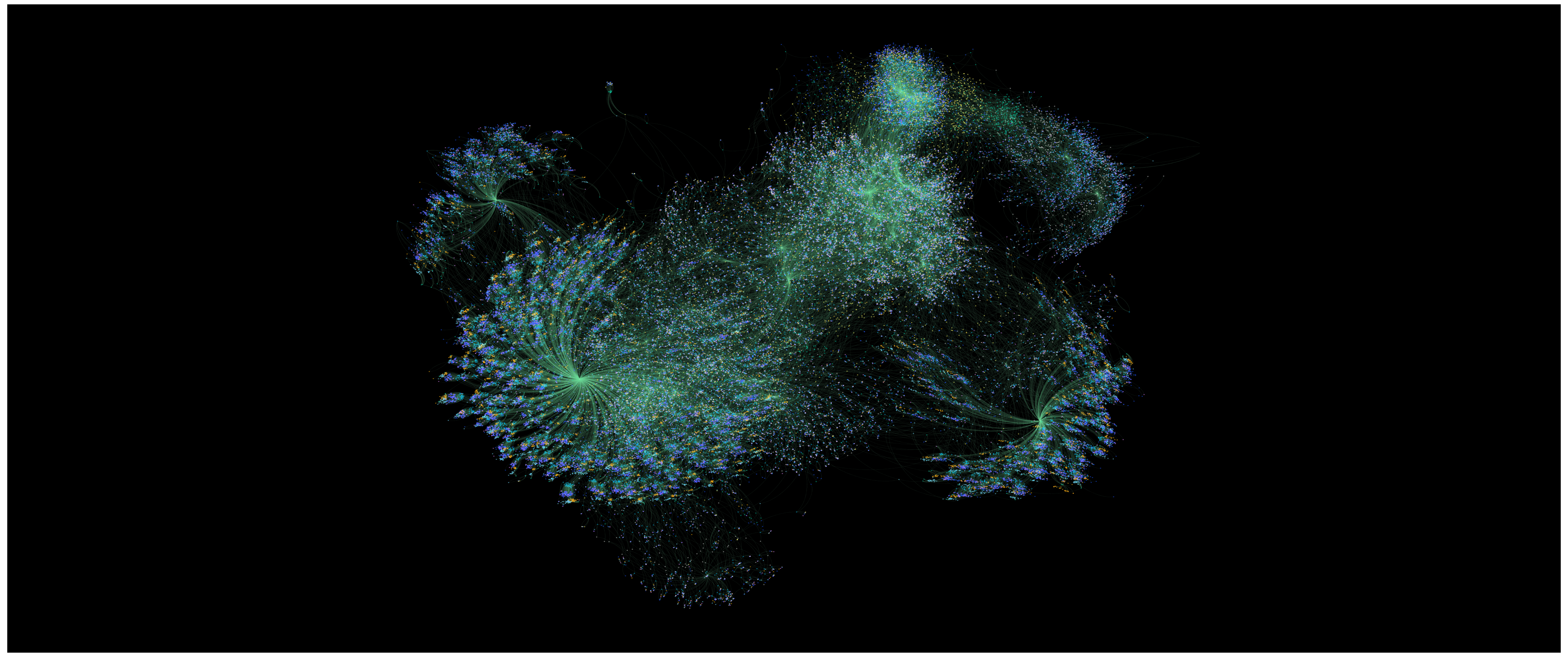
Graph requirements

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

Complexity grows quickly even for simple workflows and is impossible to reconstruct *a posteriori*



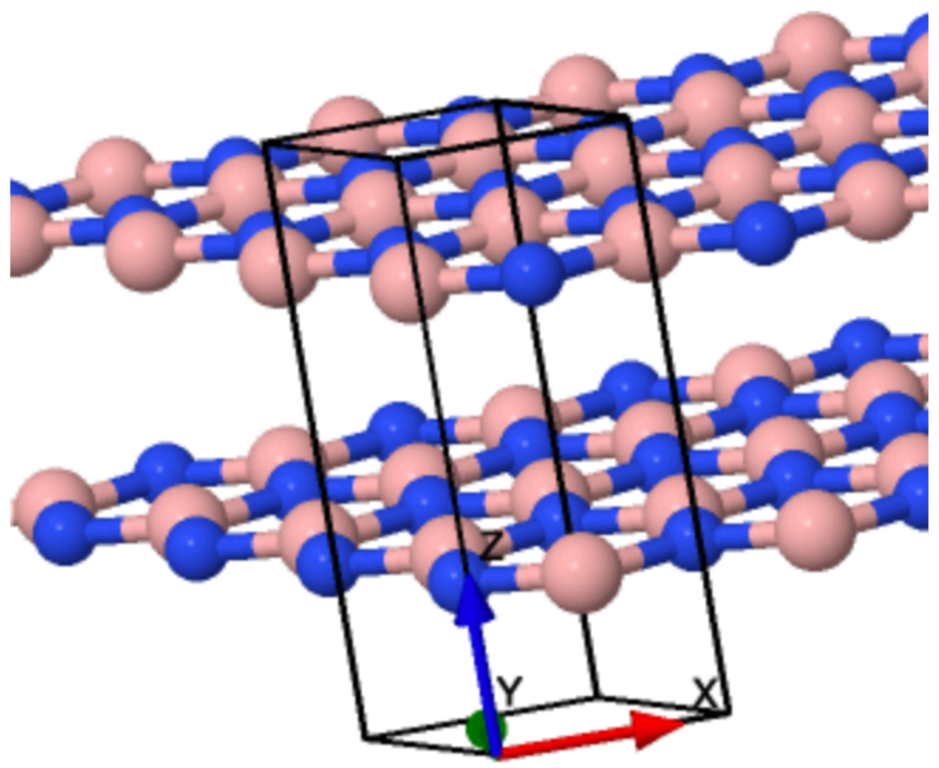
Molecular dynamics study of Lithium in a solid electrolyte

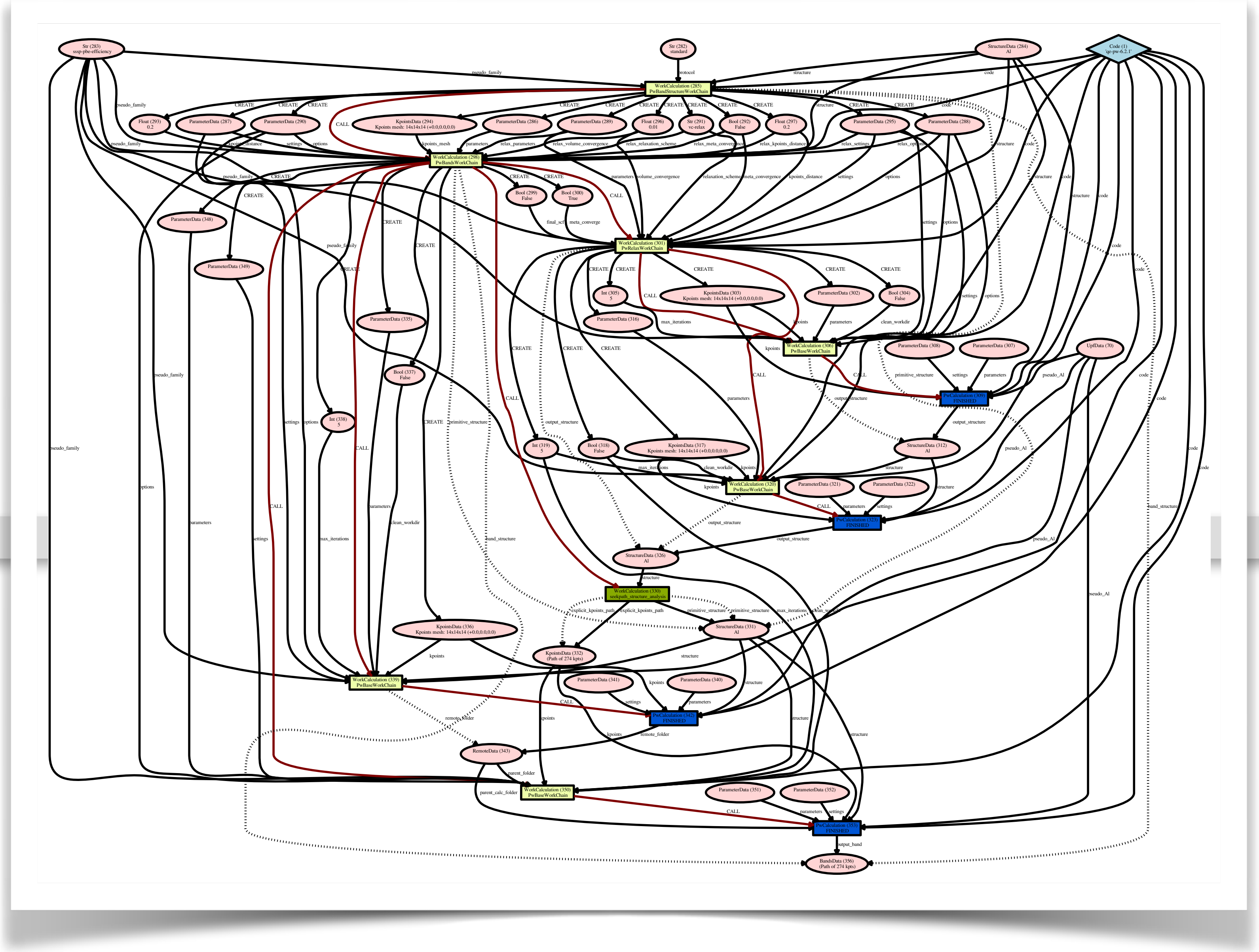


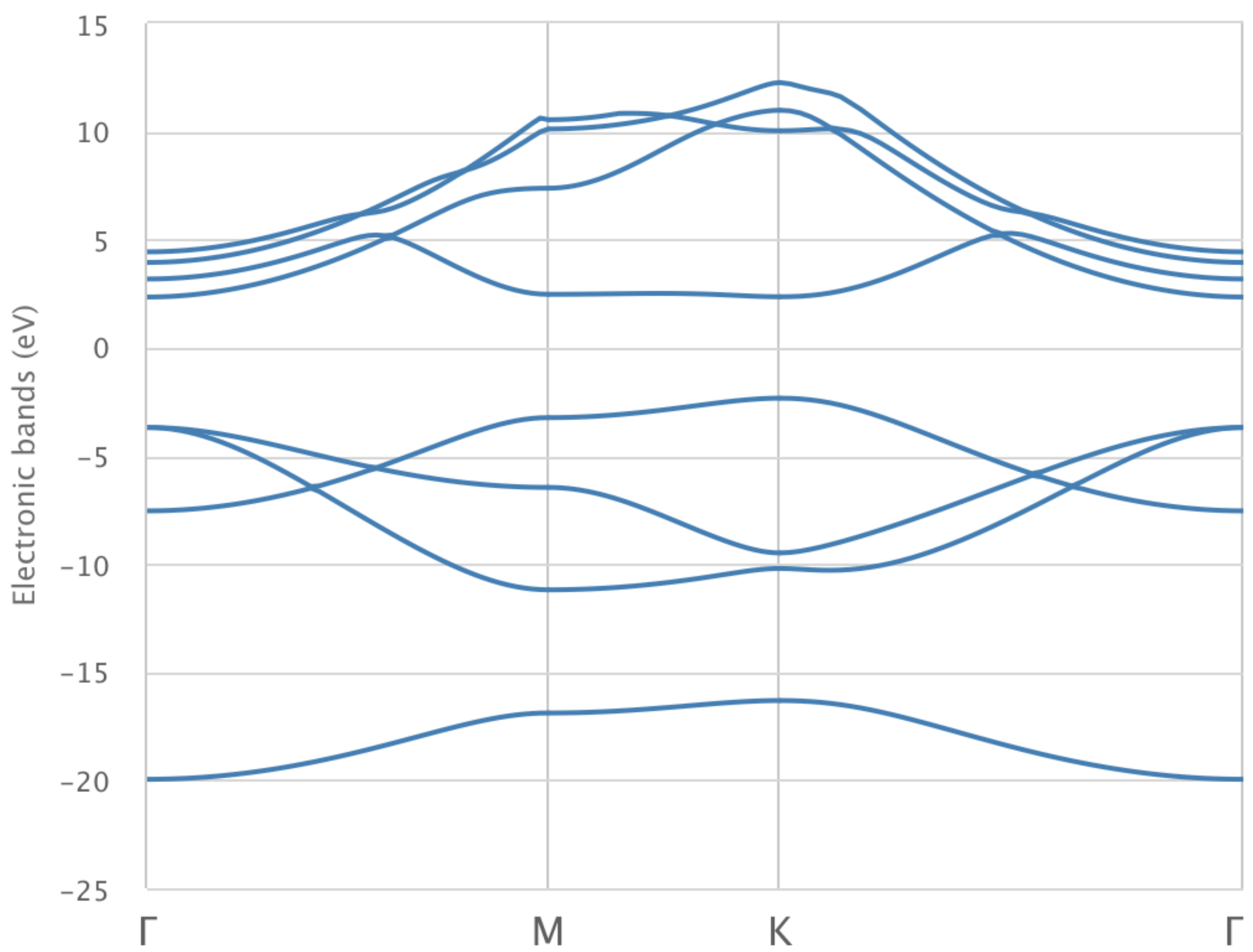
Graphical representation of actual AiiDA database

Workflows to generate data

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

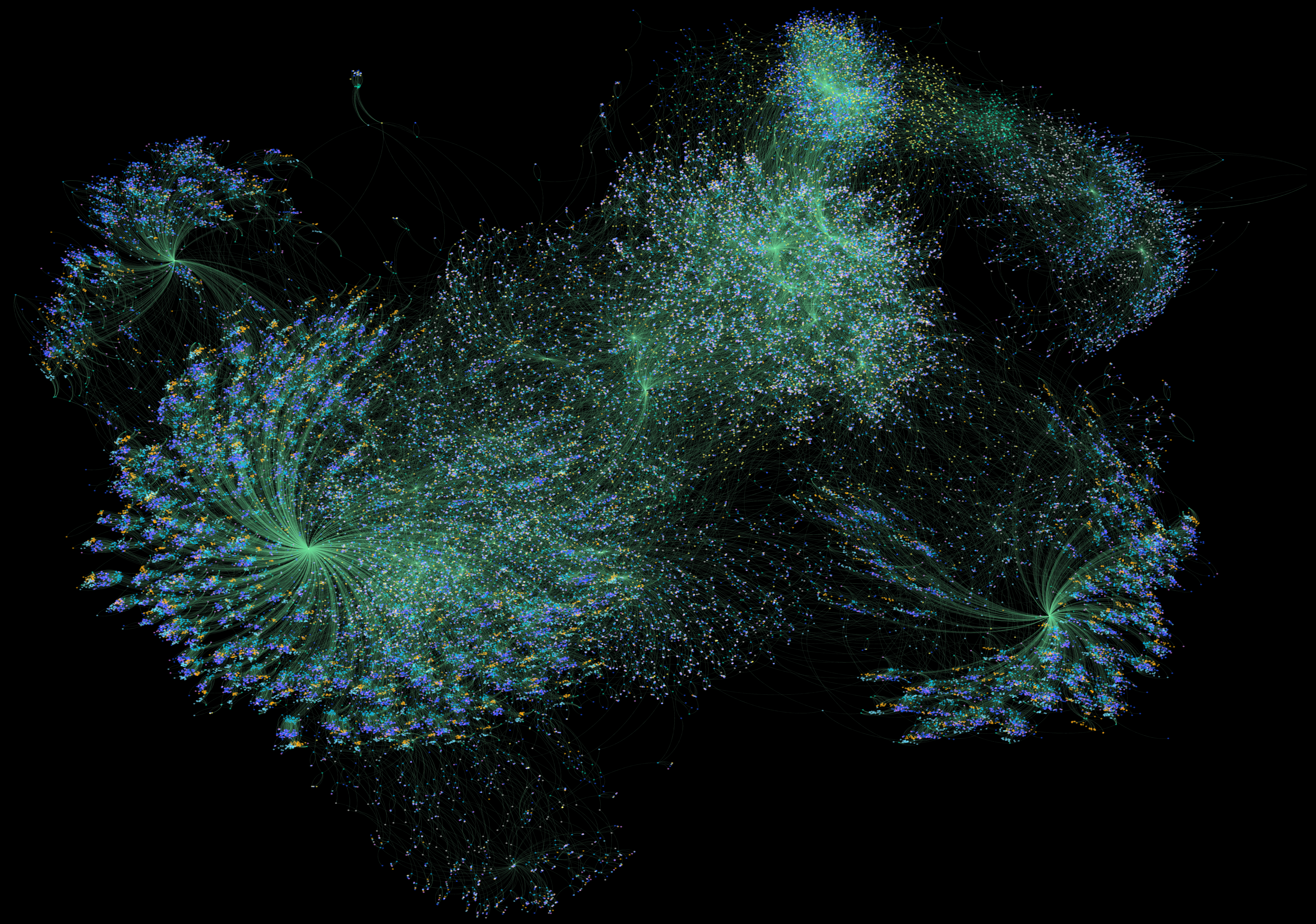






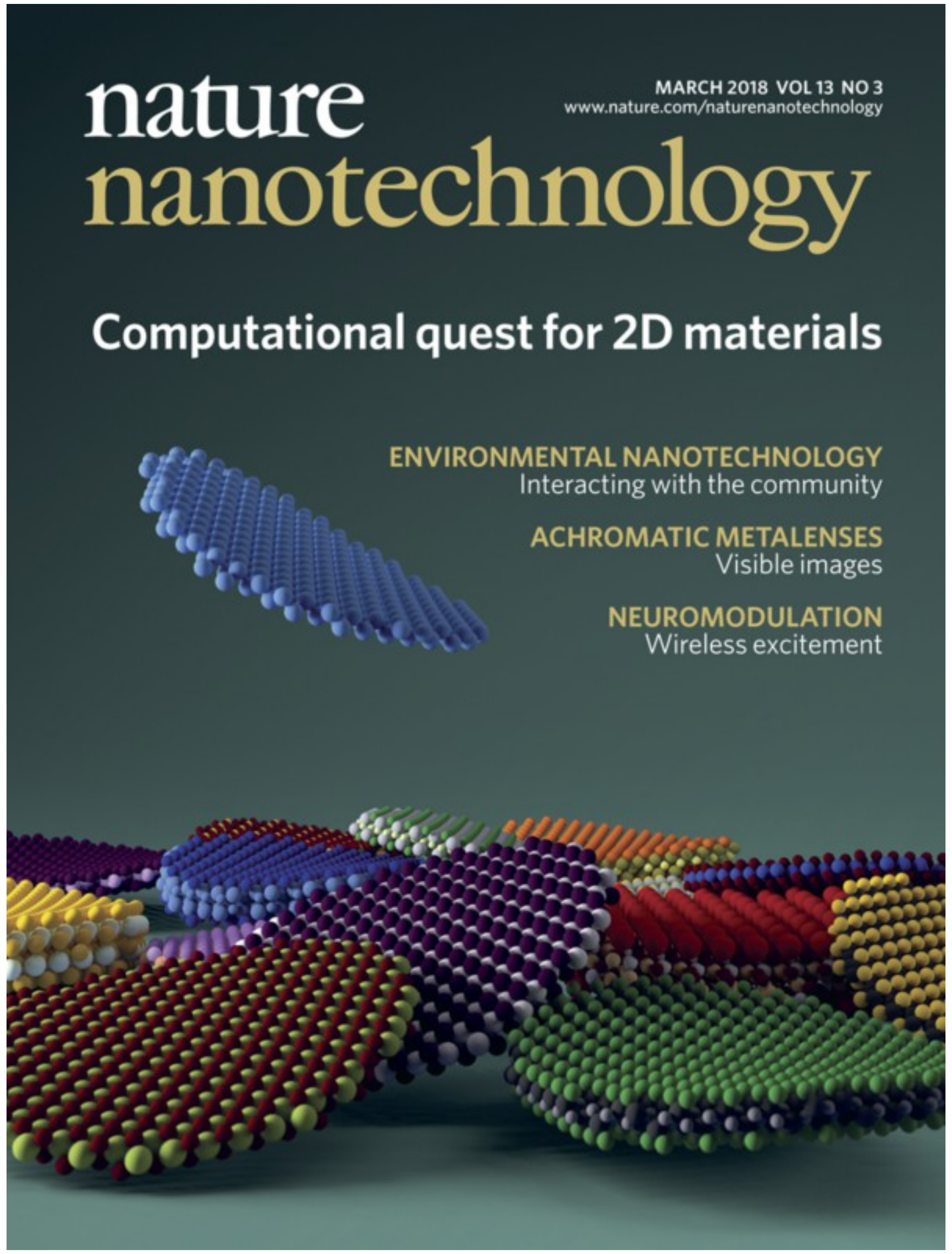
- AiiDA provenance graph: *log of "what happened in the past":*
  reproduce that single specific workflow execution
- **AiiDA workflow engine**: python interface to encode complex scientific workflows

FAIR (workflow) data sharing



Making all data open (and reproducible, and FAIR)

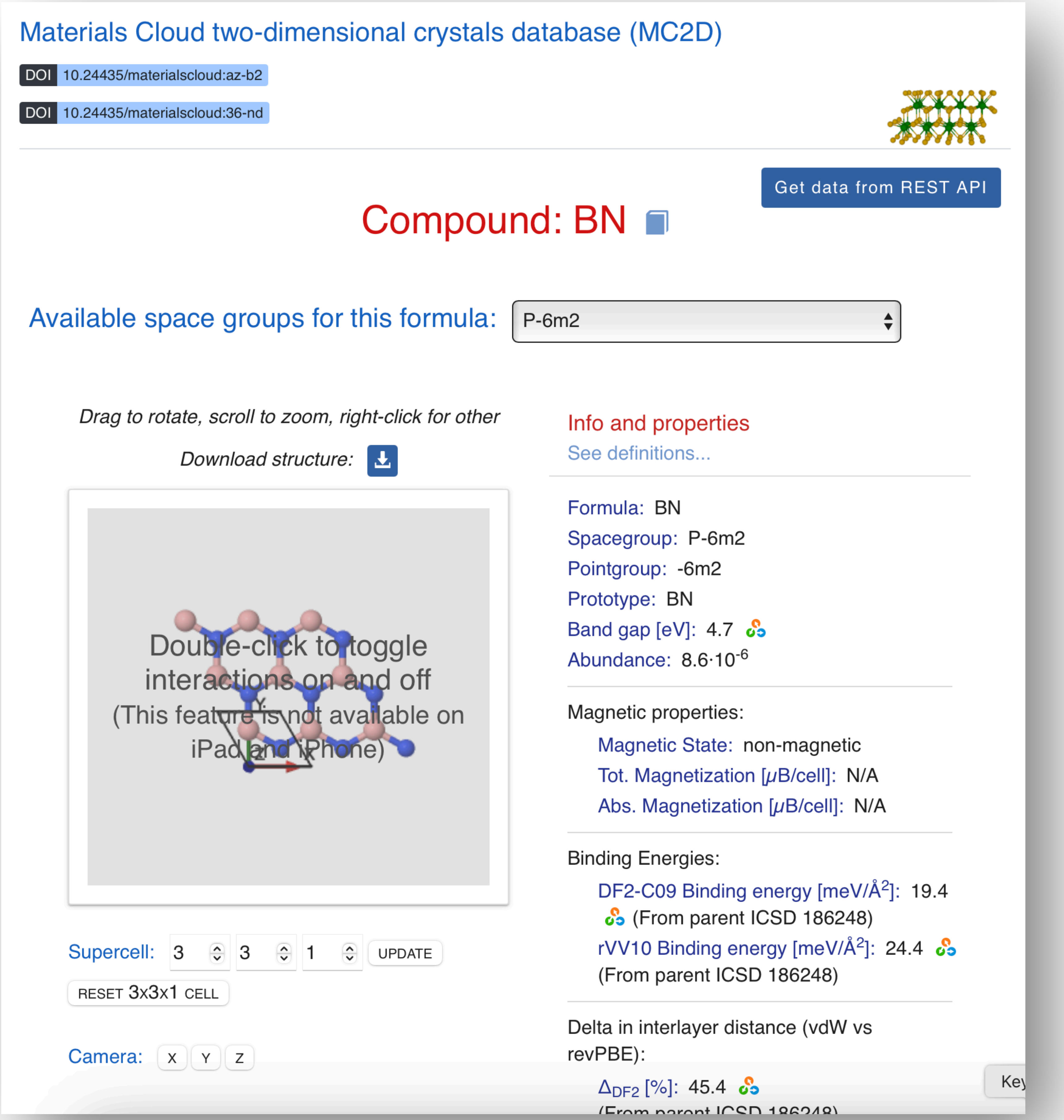
PSI



AiiDA used to obtain **MC2D, a large database of exfoliable 2D materials**

Automated workflows to perform 100'000+ DFT simulations, leading to >1800 novel 2D materials

All data open and FAIR on Materials Cloud Archive



N. Mounet et al., Nat. Nanotech. 13, 246 (2018)
D. Campi et al., ACS Nano 17 11268 (2023)

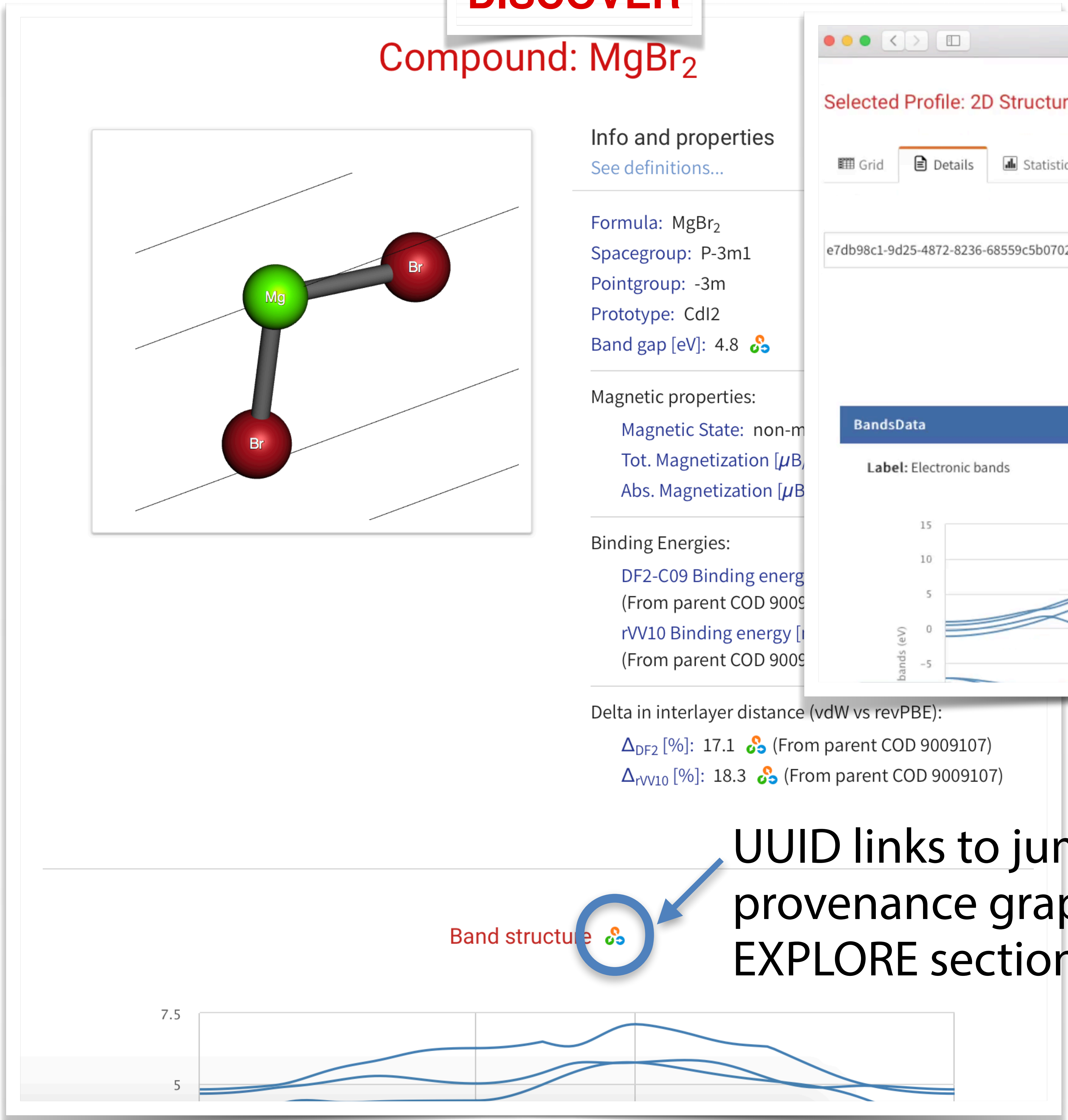
N. Mounet et al., Materials Cloud Archive 2020.158 (2020), doi: 10.24435/materialscloud:az-b2
D. Campi et al., Materials Cloud Archive 2022.84 (2022), doi: 10.24435/materialscloud:36-nd

http://mc2d.materialscloud.org

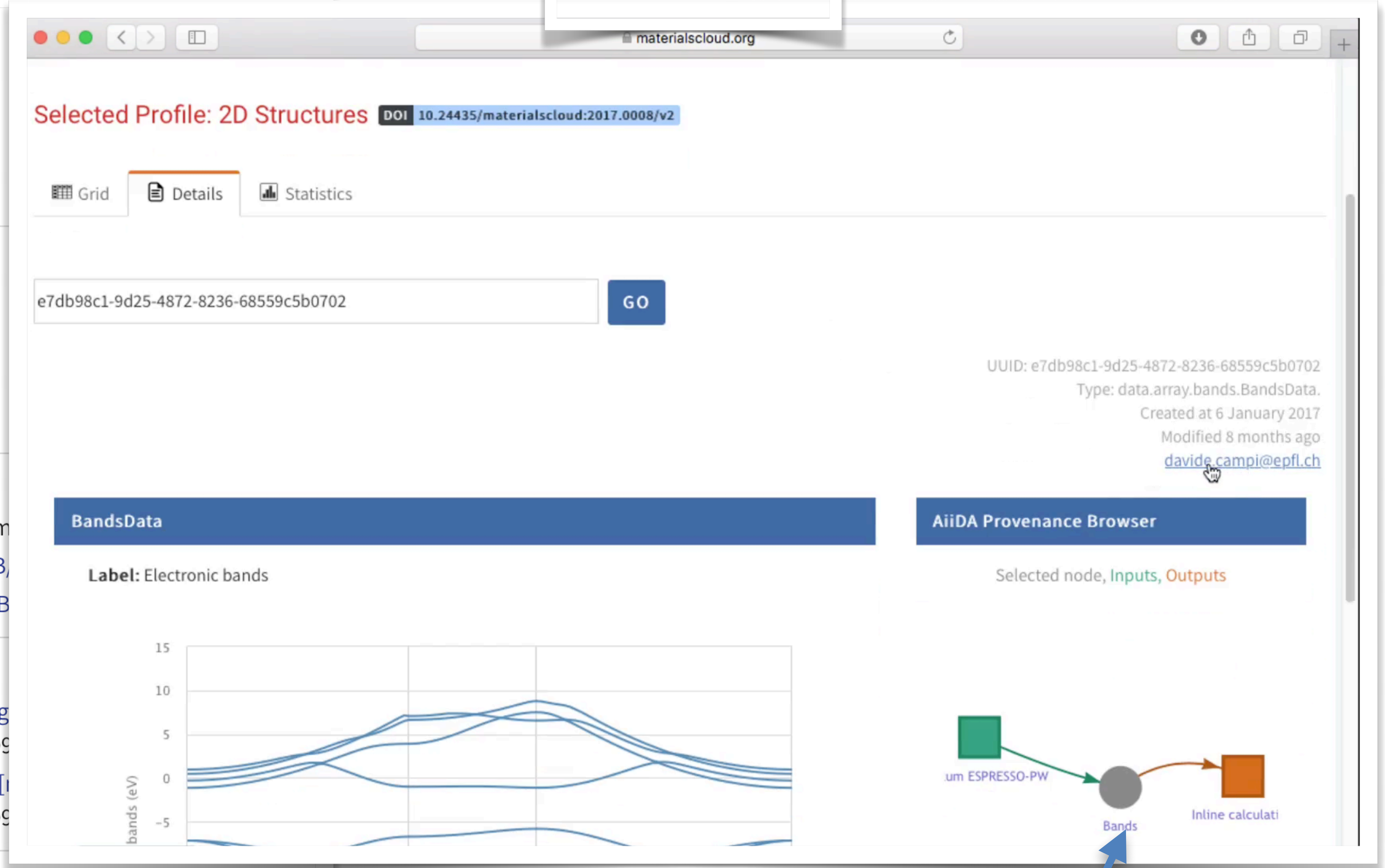
FAIR data sharing (MC2D): Materials Cloud Archive, Discover, Explore

PSI

DISCOVER



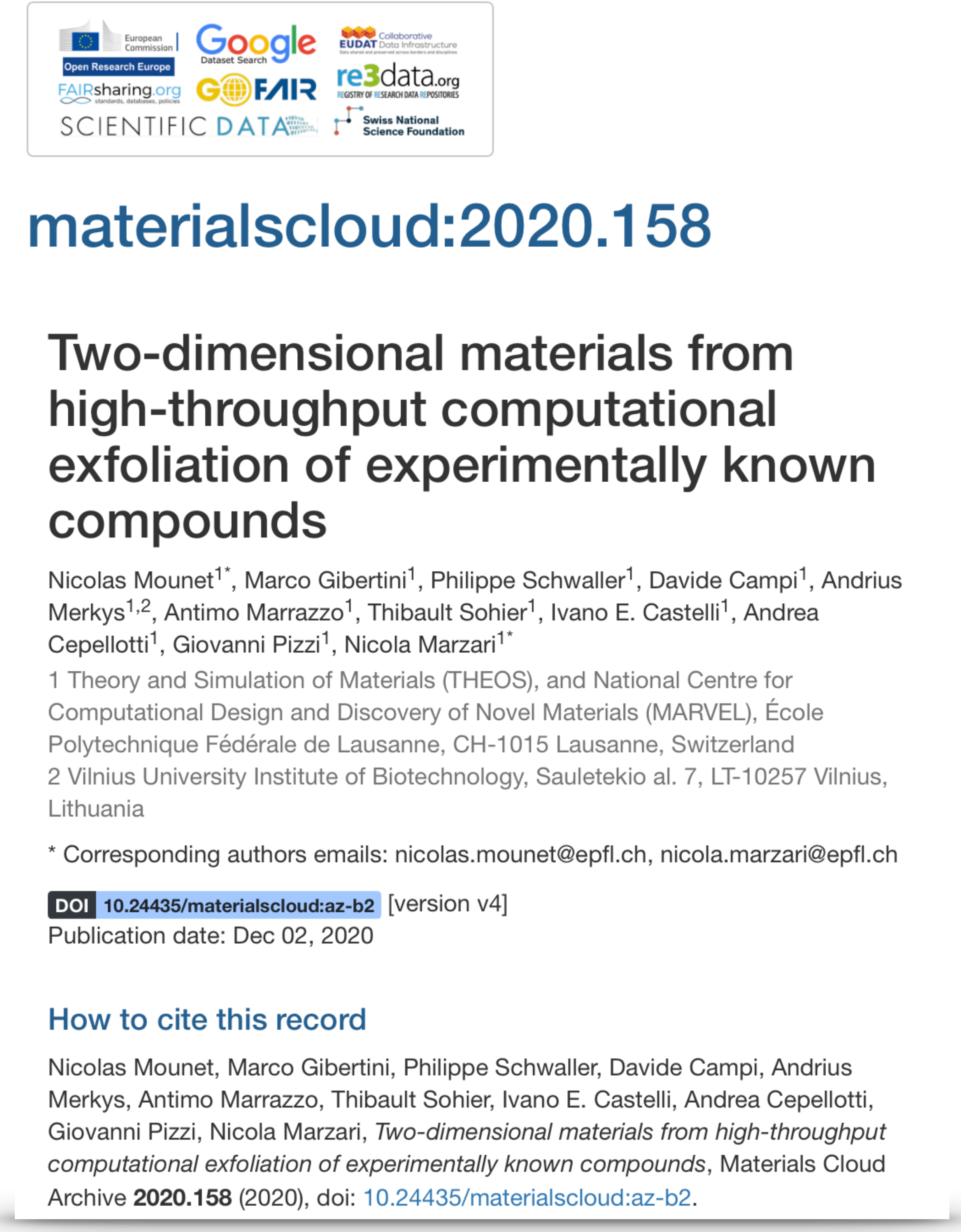
EXPLORE



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

Browse the full AiiDA provenance graph (inputs, outputs, …) at any level

ARCHIVE



N. Mounet et al., Materials Cloud Archive 2020.158 (2020), doi: 10.24435/materialscloud:az-b2

http://mc2d.materialscloud.org

Research data repository: Materials Cloud Archive

PSI

DOIs assigned

Direct links to Discover & Explore

Data (and metadata) guaranteed to be online for at least 10 years after deposition

materialscloud:2017.0008/v3

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

Nicolas Mounet[1*], Marco Gibertini[1], Philippe Schwaller[1], Davide Campi[1], Andrius Merkys[1,2], Antimo Marrazzo[1], Thibault Sohier[1], Ivano E. Castelli[1], Andrea Cepellotti[1], Giovanni Pizzi[1], Nicola Marzari[1*]

1 Theory and Simulation of Materials (THEOS), and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

2 Vilnius University Institute of Biotechnology, Sauletekio al. 7, LT-10257 Vilnius, Lithuania

* Corresponding authors emails: nicolas.mounet@epfl.ch, nicola.marzari@epfl.ch

DOI 10.24435/materialscloud:2017.0008/v3 [version v3]
Publication date: Apr 03, 2019

How to cite this record

Nicolas Mounet, Marco Gibertini, Philippe Schwaller, Davide Campi, Andrius Merkys, Antimo Marrazzo, Thibault Sohier, Ivano E. Castelli, Andrea Cepellotti, Giovanni Pizzi, Nicola Marzari, *Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds*, Materials Cloud Archive **2017.0008/v3** (2019), doi: 10.24435/materialscloud:2017.0008/v3.

Description

Two-dimensional (2D) materials have emerged as promising candidates for next-generation electronic and optoelectronic applications. Yet, only a few dozens of 2D materials have been successfully synthesized or exfoliated. Here, we search for novel 2D materials that can be easily exfoliated from their parent compounds. Starting from 108423 unique, experimentally known three-dimensional compounds we identify a subset of 5619 that appear layered according to robust geometric and bonding criteria. High-throughput calculations using van-der-Waals density-functional theory, validated against experimental structural data and calculated random-phase-approximation binding energies, allow to identify 1825 compounds that are either easily or potentially exfoliable. In particular, the subset of 1036 easily exfoliable cases provides novel structural prototypes and simple ternary compounds as well as a large portfolio of materials to search from for optimal properties. For a subset of 258 compounds we explore vibrational, electronic, magnetic, and topological properties, identifying 56 ferromagnetic and antiferromagnetic systems, including half-metals and half-semiconductors. This archive entry contains the database of 2D materials (structural parameters, band structures, binding energies, phonons for the subset of the 258 easily exfoliable materials with less than 6 atoms, structures and binding energies for the remaining 1567 materials) together with the provenance of all data and calculations as stored by AiiDA.

Materials Cloud sections using this data

- Select 2d materials via interactive periodic table and view their properties (with links to provenance)
- Explore interface providing access to the full database

Files

File name	Size	Description
2D_materials.tar.gz MD5	113.0 MiB	We provide 258 two-dimensional crystal structures (lattice vectors, atomic species and positions), exfoliated from three-dimensional experimental crystal structures. The structures were relaxed at the DFT-PBE level. Together with each structure, a set of materials properties is also given (at the DFT-PBE level): chemical formula, spacegroup, structural prototype, magnetic state, magnetization, band-gap, electronic bands, and phonon

Export
Dublin Core JSON

Recommended repository by Nature's journal Scientific Data, the EU Open Research Europe , and SNSF

SCIENTIFIC DATA

European Commission

Research and Innovation

Open Research Europe

Indexed by Google Dataset Search and EUDAT/EOSC's B2FIND; Registered on FAIRsharing.org and re3data.org

Currently using (extended version of) **CERN's Invenio v3**

Now **migrating to InvenioRDM v12**

Accessible simulation capabilities

PSI

FAIR sharing in AiiDA beyond data: codes, plugins and workflows

PSI

Calculation | **Data** | **Parsers** | **Transport and scheduler** | **Workflows** | **Importers & exporters**

AiiDA PLUGIN REGISTRY

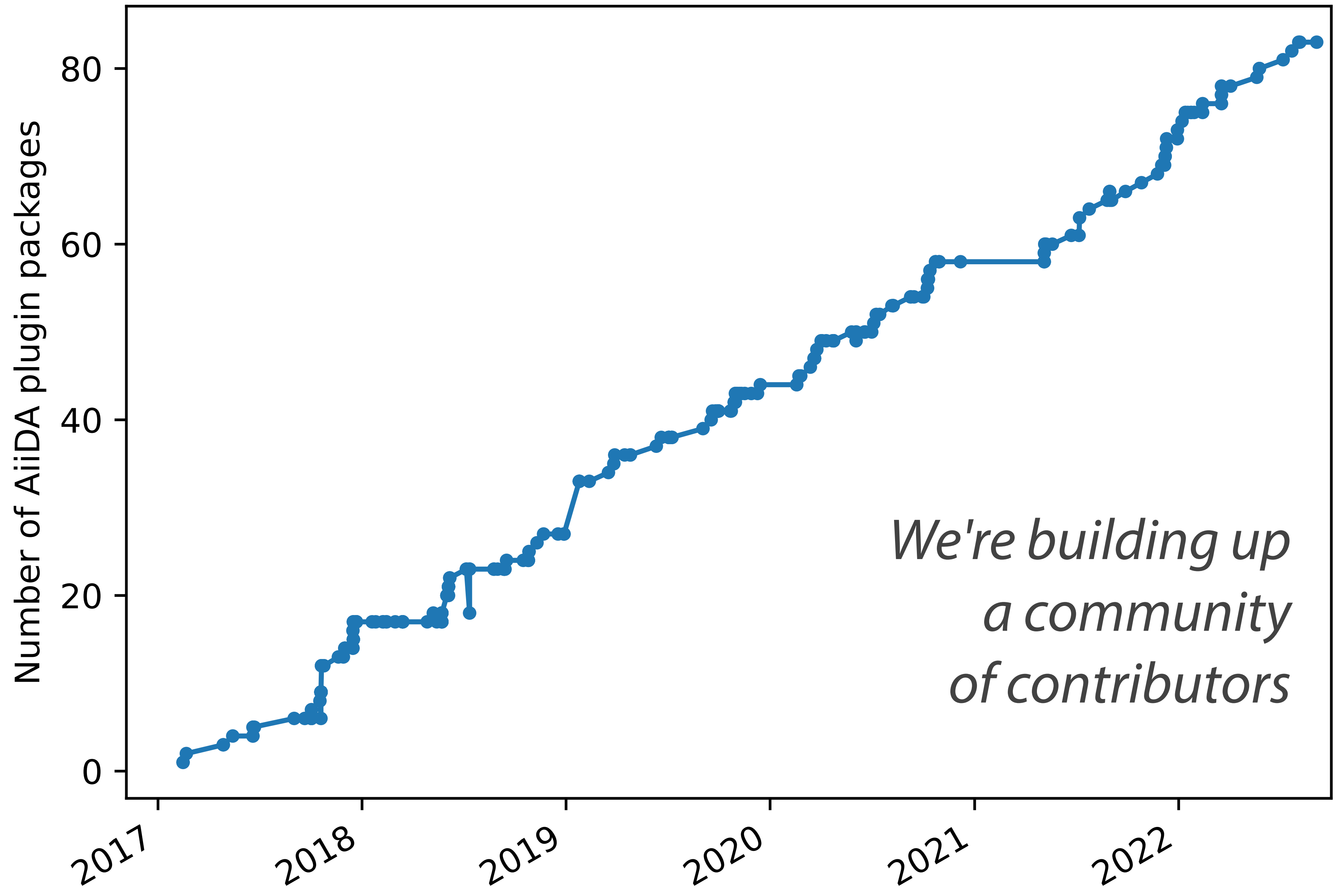
[View on GitHub/register your package]

Registered plugin packages: 97

Calculations	164 plugins in 66 packages
Parsers	145 plugins in 67 packages
Data	122 plugins in 38 packages
Workflows	211 plugins in 50 packages
Console scripts	33 plugins in 17 packages
Other	110 plugins in 33 packages

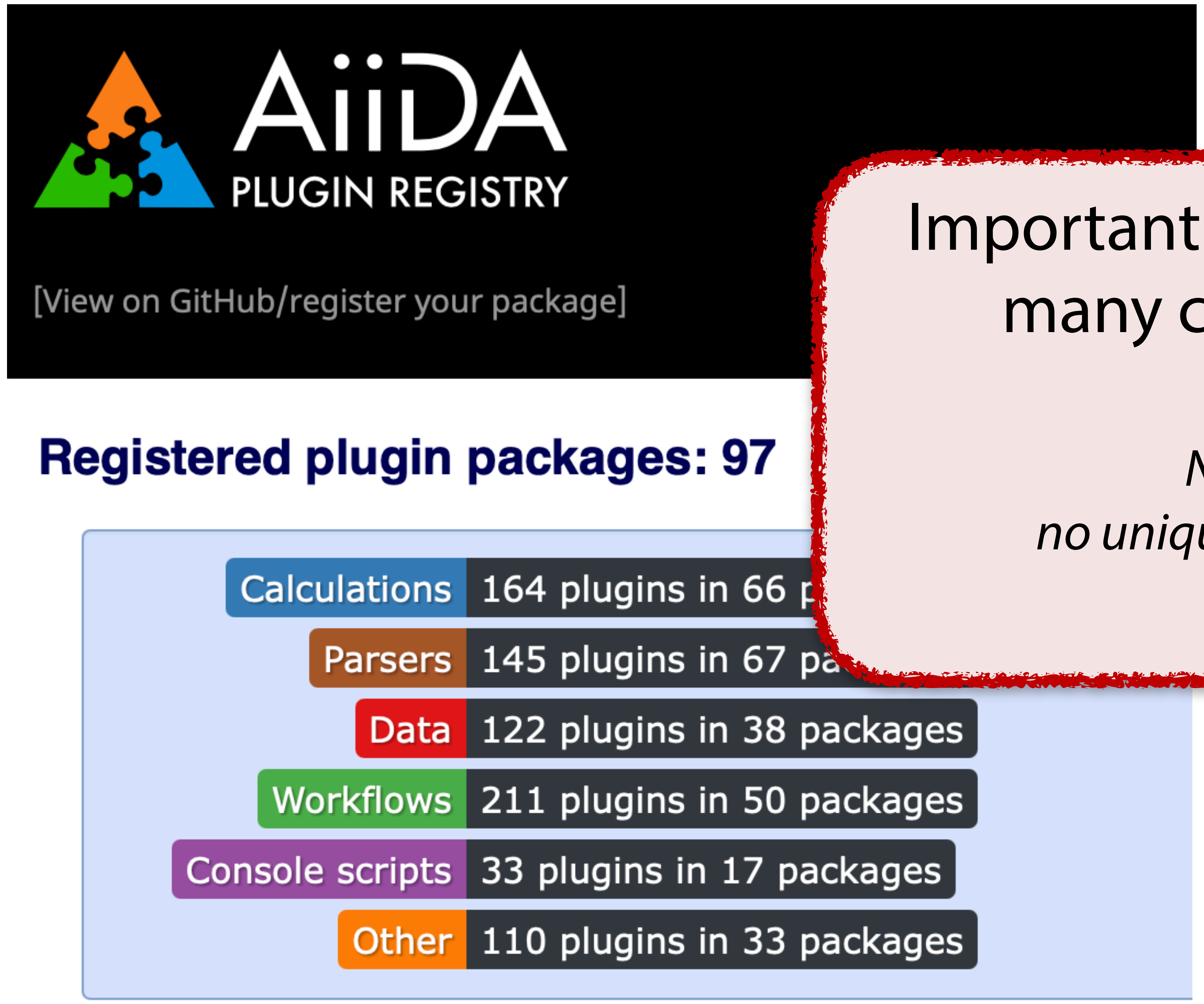
https://aiidateam.github.io/aiida-registry/

- Plugins collected in the AiiDA plugin registry
- **160+ codes currently supported,** 200+ workflows
- Many are **community-contributed**



FAIR sharing in AiiDA beyond data: codes, plugins and workflows

Calculation | **Data** | **Parsers** | **Transport and scheduler** | **Workflows** | **Importers & exporters**

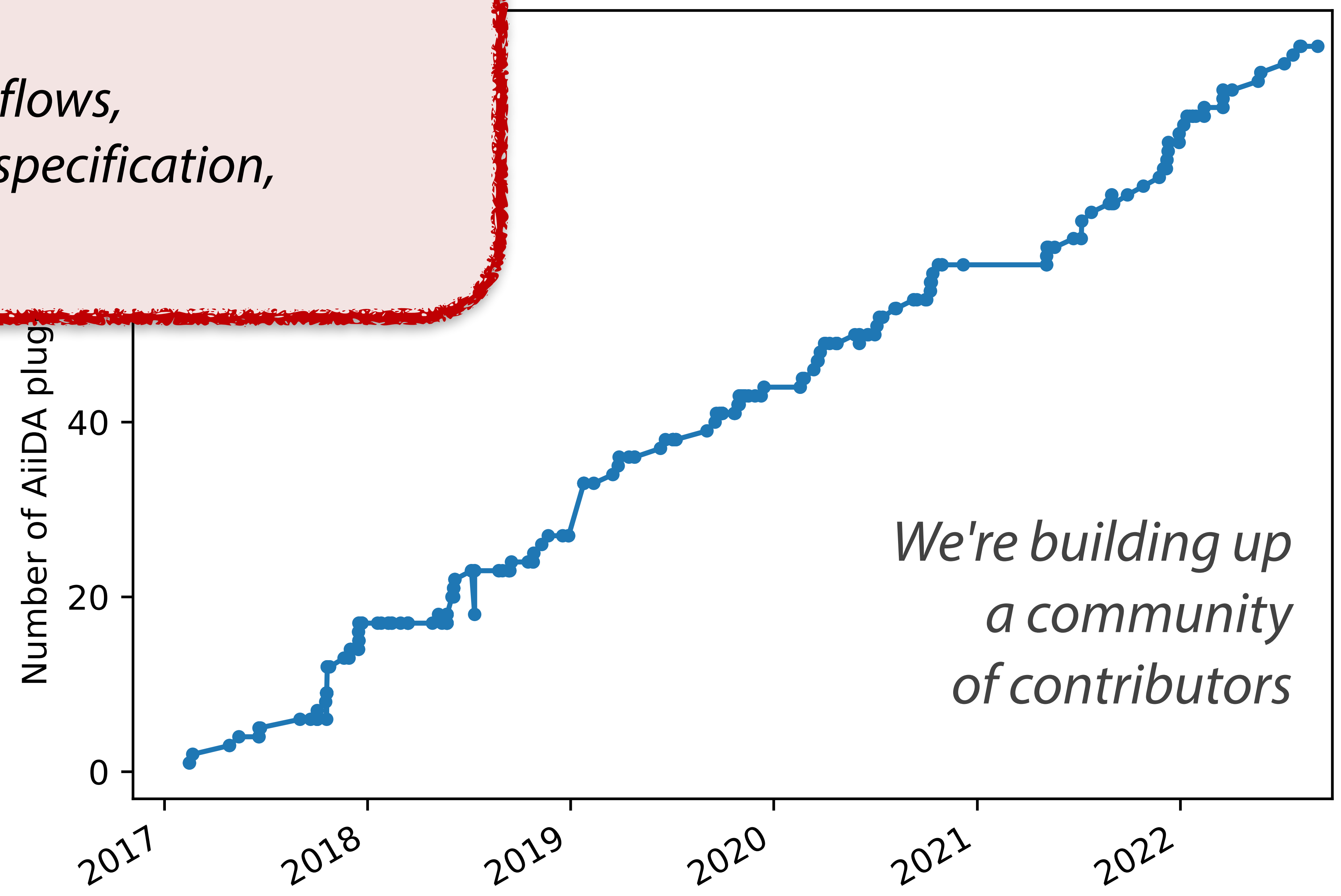


https://aiidateam.github.io/aiida-registry/

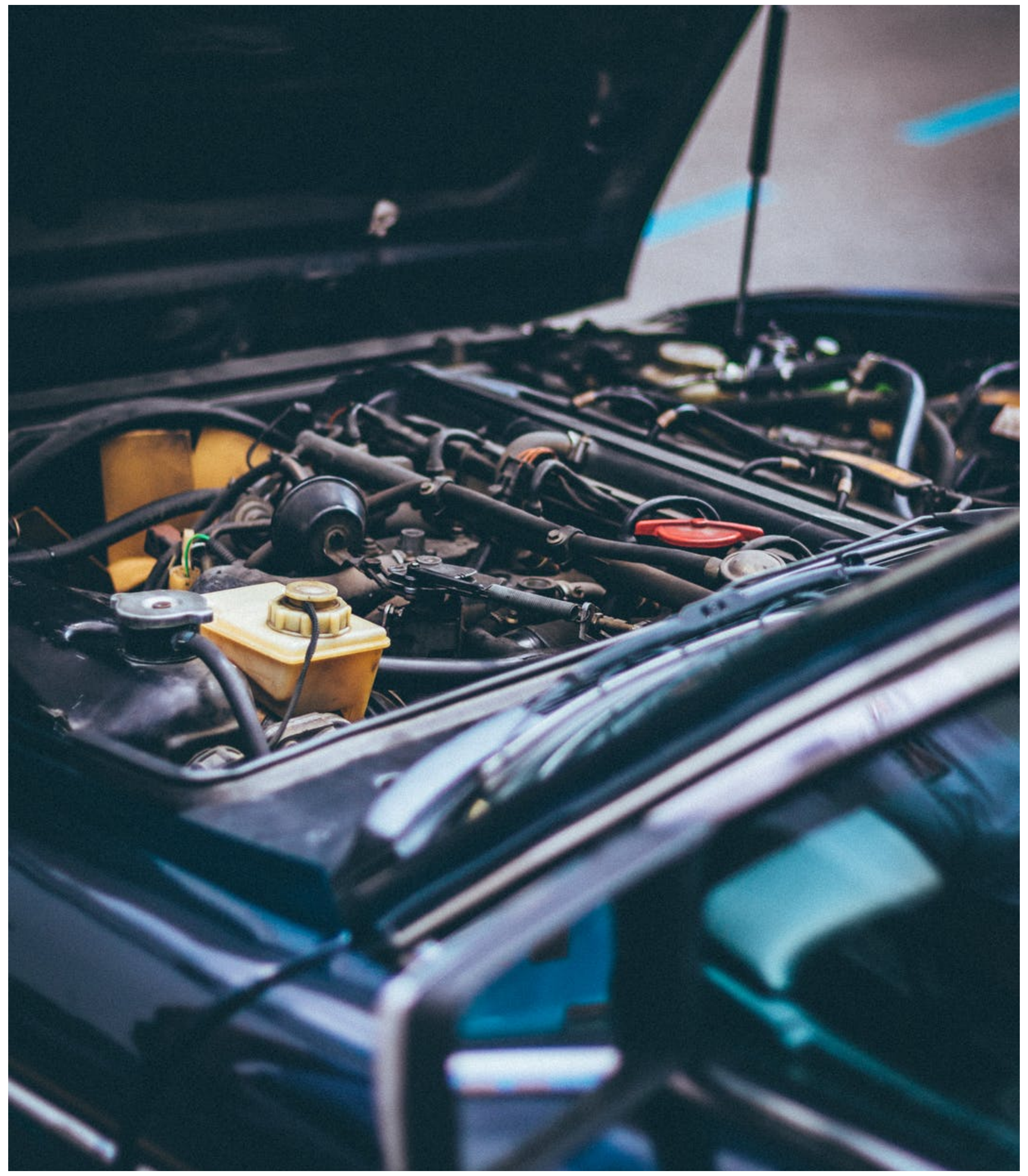
- Plugins collected in the AiiDA plugin registry
  …tly supported, 200+ workflows
  …y-contributed

Important "requirement" to support many codes, but not enough!

No turn-key workflows,
no unique input/output specification,
…



The need for turn-key solutions



Like in a car: drive without needing to know how the engine works

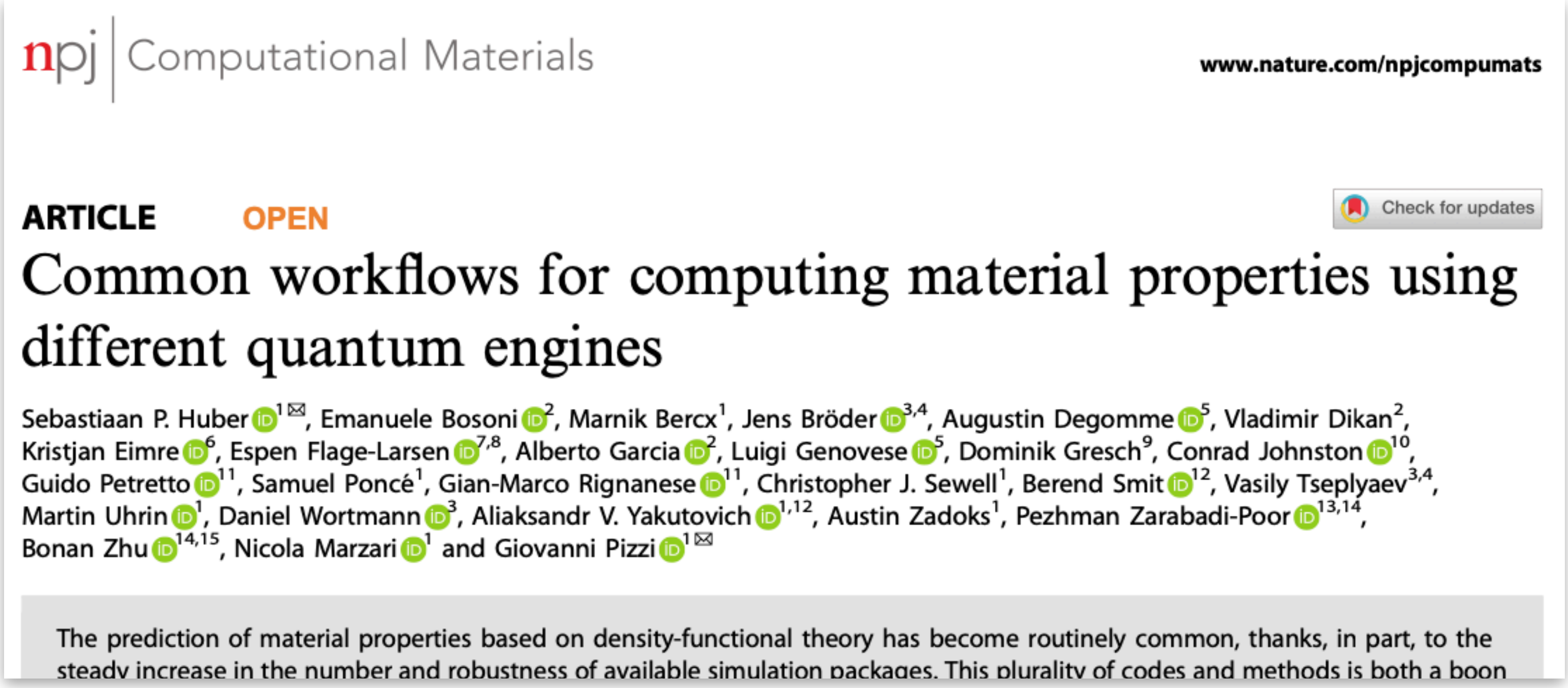
- Engines are "robust"
- Just turn the key and drive
- I need a driving license, but I don't need to learn again if I change the brand of my car

Interoperable workflows:
AiiDA common workflow interfaces (ACWF)

As a non-expert, be able to ask

"Please run an **equation of state** with code **[Quantum ESPRESSO|SIESTA|VASP|...]** on the XXX supercomputer, using YY nodes, and **automatically choose numerical parameters** to get converged results."

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


npj | Computational Materials www.nature.com/npjcompumats

ARTICLE OPEN Check for updates

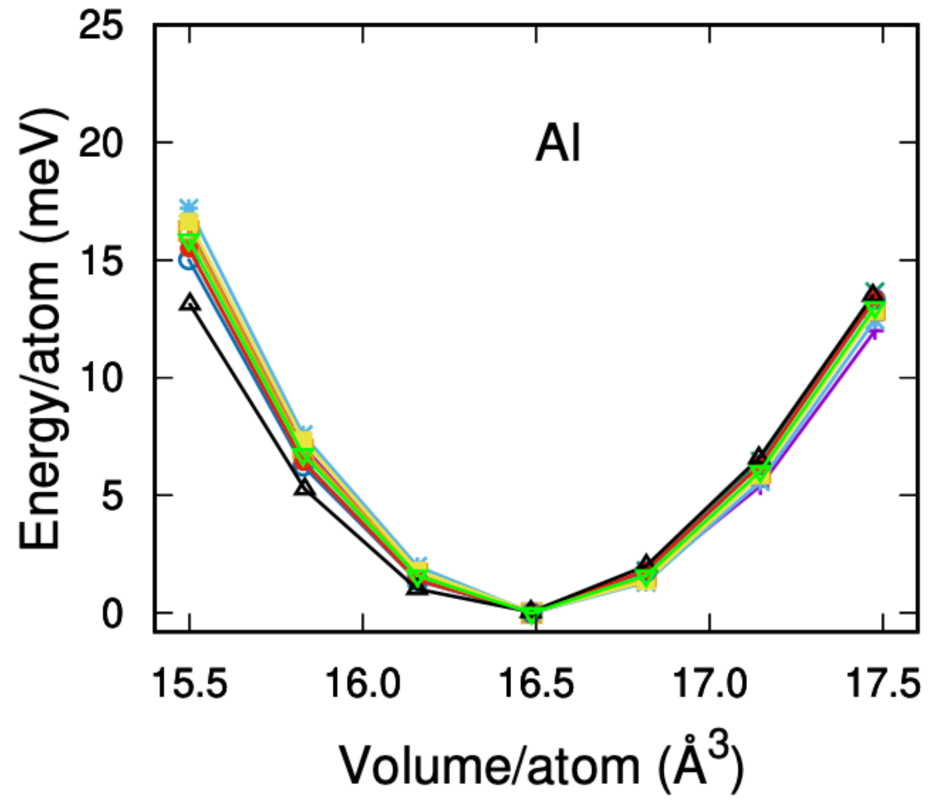
Common workflows for computing material properties using different quantum engines

Sebastiaan P. Huber[1 ✉], Emanuele Bosoni[2], Marnik Bercx[1], Jens Bröder[3,4], Augustin Degomme[5], Vladimir Dikan[2], Kristjan Eimre[6], Espen Flage-Larsen[7,8], Alberto Garcia[2], Luigi Genovese[5], Dominik Gresch[9], Conrad Johnston[10], Guido Petretto[11], Samuel Poncé[1], Gian-Marco Rignanese[11], Christopher J. Sewell[1], Berend Smit[12], Vasily Tseplyaev[3,4], Martin Uhrin[1], Daniel Wortmann[3], Aliaksandr V. Yakutovich[1,12], Austin Zadoks[1], Pezhman Zarabadi-Poor[13,14], Bonan Zhu[14,15], Nicola Marzari[1] and Giovanni Pizzi[1 ✉]

The prediction of material properties based on density-functional theory has become routinely common, thanks, in part, to the steady increase in the number and robustness of available simulation packages. This plurality of codes and methods is both a boon

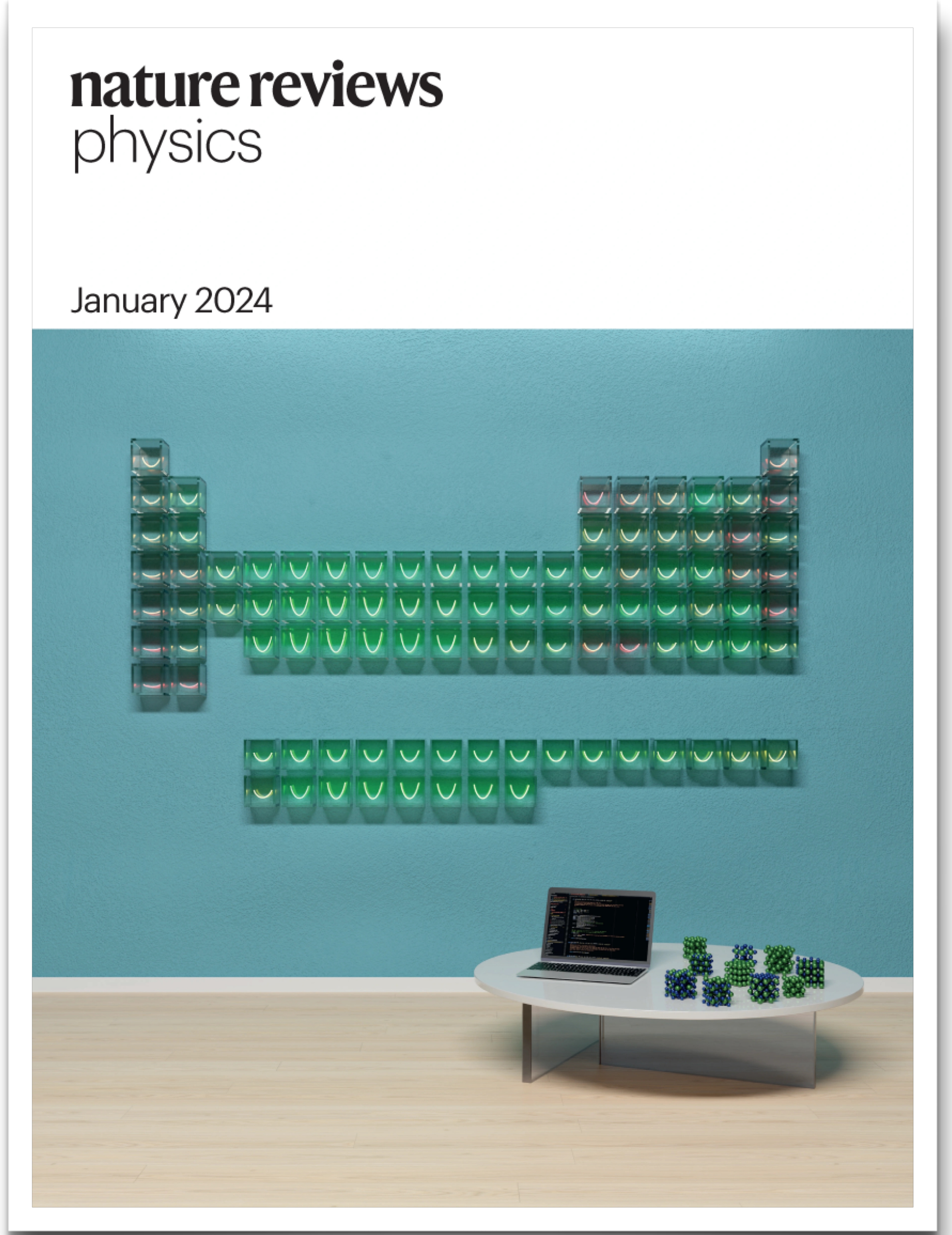
S. P. Huber et al., npj Comput. Mater. 7, 136 (2021)

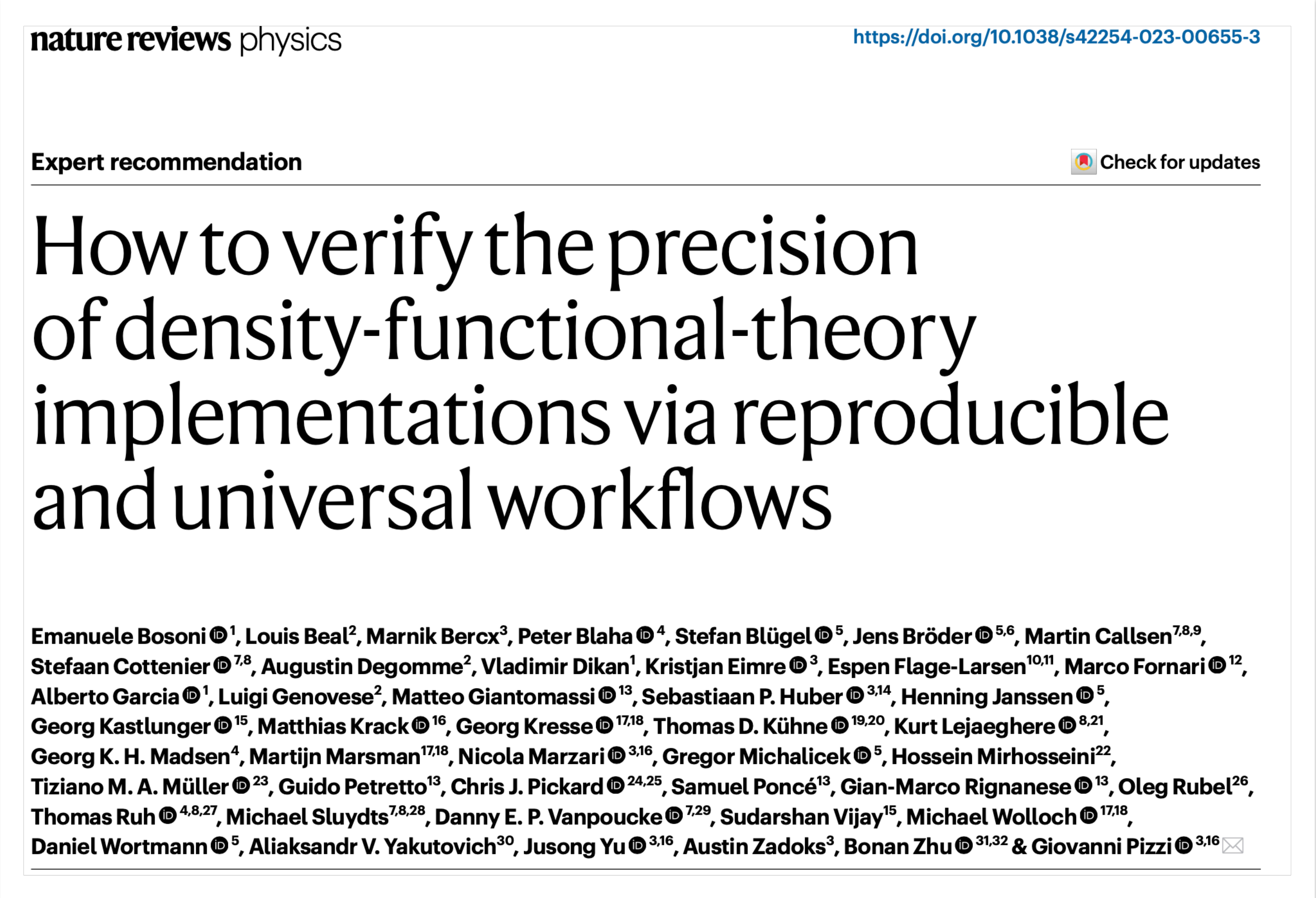
$ aiida-common-workflows launch **eos siesta** --structure=**Al** --protocol=**precise**



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

Enabling code verification via AiiDA common workflows




nature reviews physics https://doi.org/10.1038/s42254-023-00655-3

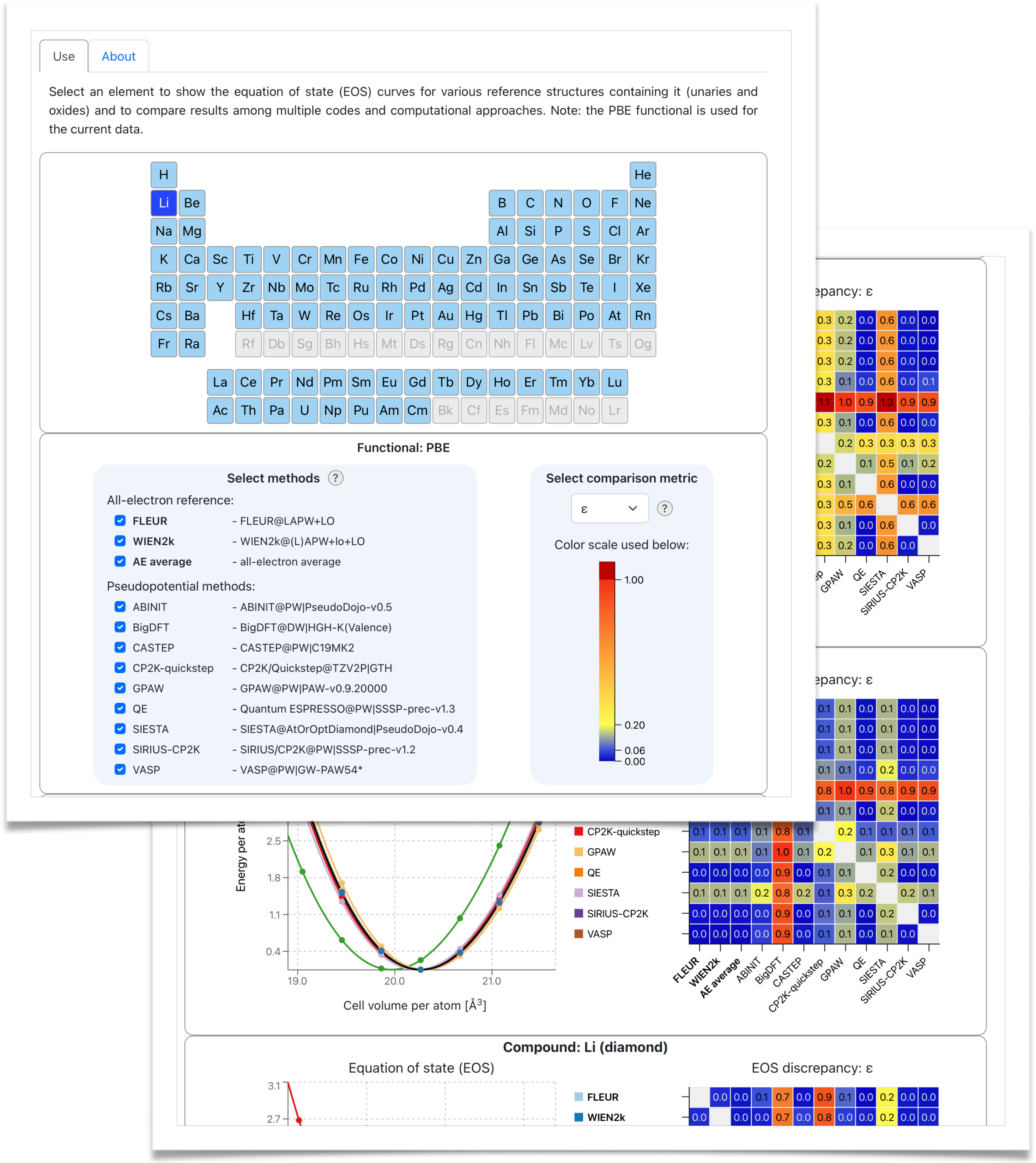
Expert recommendation

Check for updates

How to verify the precision of density-functional-theory implementations via reproducible and universal workflows

Emanuele Bosoni[1], Louis Beal[2], Marnik Bercx[3], Peter Blaha[4], Stefan Blügel[5], Jens Bröder[5,6], Martin Callsen[7,8,9], Stefaan Cottenier[7,8], Augustin Degomme[2], Vladimir Dikan[1], Kristjan Eimre[3], Espen Flage-Larsen[10,11], Marco Fornari[12], Alberto Garcia[1], Luigi Genovese[2], Matteo Giantomassi[13], Sebastiaan P. Huber[3,14], Henning Janssen[5], Georg Kastlunger[15], Matthias Krack[16], Georg Kresse[17,18], Thomas D. Kühne[19,20], Kurt Lejaeghere[8,21], Georg K. H. Madsen[4], Martijn Marsman[17,18], Nicola Marzari[3,16], Gregor Michalicek[5], Hossein Mirhosseini[22], Tiziano M. A. Müller[23], Guido Petretto[13], Chris J. Pickard[24,25], Samuel Poncé[13], Gian-Marco Rignanese[13], Oleg Rubel[26], Thomas Ruh[4,8,27], Michael Sluydts[7,8,28], Danny E. P. Vanpoucke[7,29], Sudarshan Vijay[15], Michael Wolloch[17,18], Daniel Wortmann[5], Aliaksandr V. Yakutovich[30], Jusong Yu[3,16], Austin Zadoks[3], Bonan Zhu[31,32] & Giovanni Pizzi[3,16]

E. Bosoni *et al.*, Nat. Rev. Phys. 6, 45 (2024)



http://acwf-verification.materialscloud.org

Comparison of 11 codes and computational approaches
(algorithms, basis sets, pseudopotentials, ...)

Robustness is important (for code users)

Many efforts in the community: **more efficient, faster codes, new architectures** (GPUs, ...)

What about robustness? (guarantee to converge to solution): crucial, but (probably?) gets less attention

Robustness is important (for code users)

PSI

Many efforts in the community: **more efficient, faster codes, new architectures** (GPUs, ...)

What about robustness? (guarantee to converge to solution): crucial, but (probably?) gets less attention

Question: Would you buy a car that:

- costs 1/3; consumes 1/3;
- but randomly stops in the middle of the road 1/3 of your trips? (Every other day!)

Making simulations accessible to all

PSI

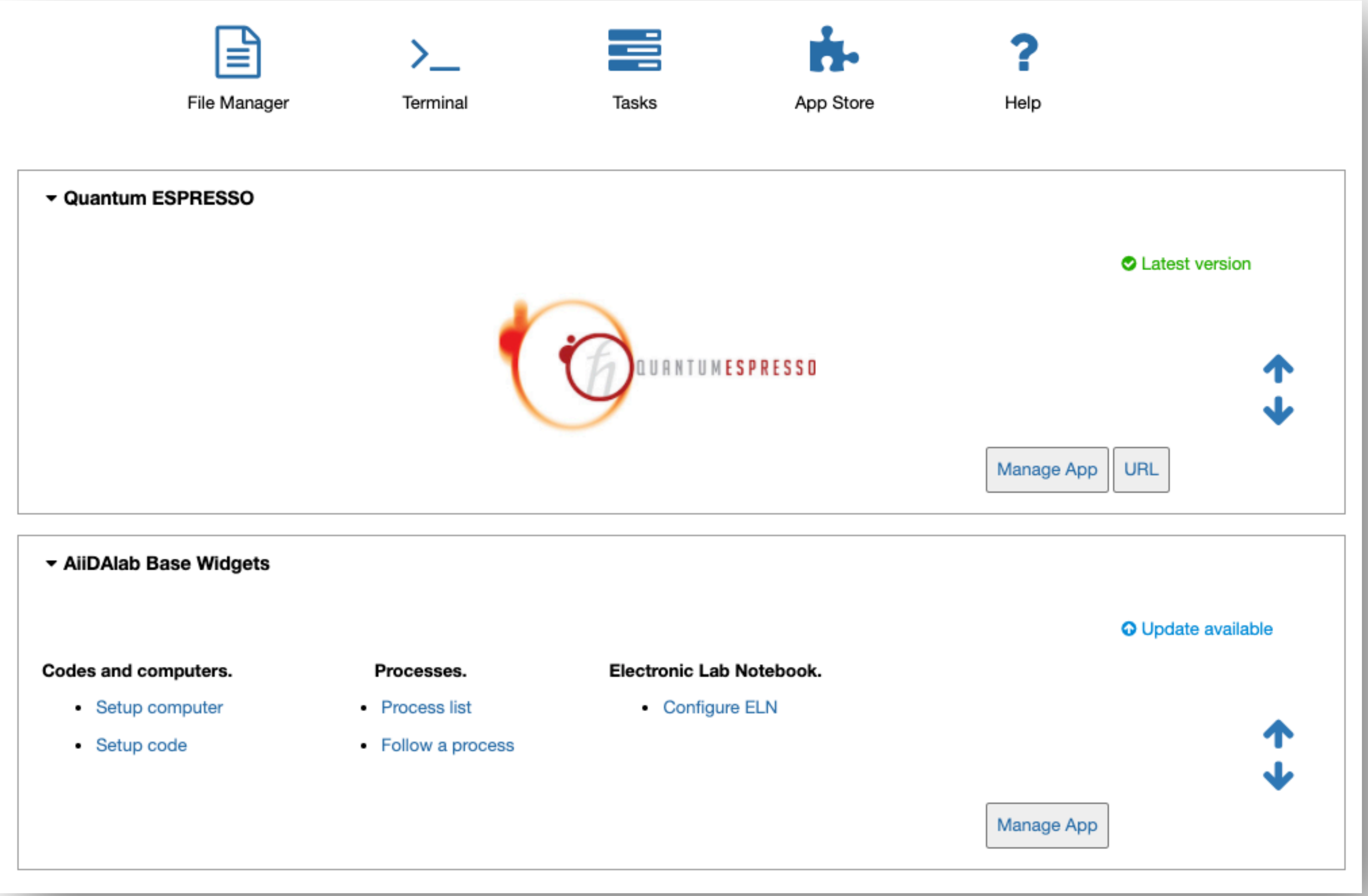
AiiDAlab

https://www.aiidalab.net

A. V. Yakutovich et al., Comp. Mat. Sci. 188, 110165 (2021)

AiiDAlab: easy access to simulation capabilities to everybody

PSI

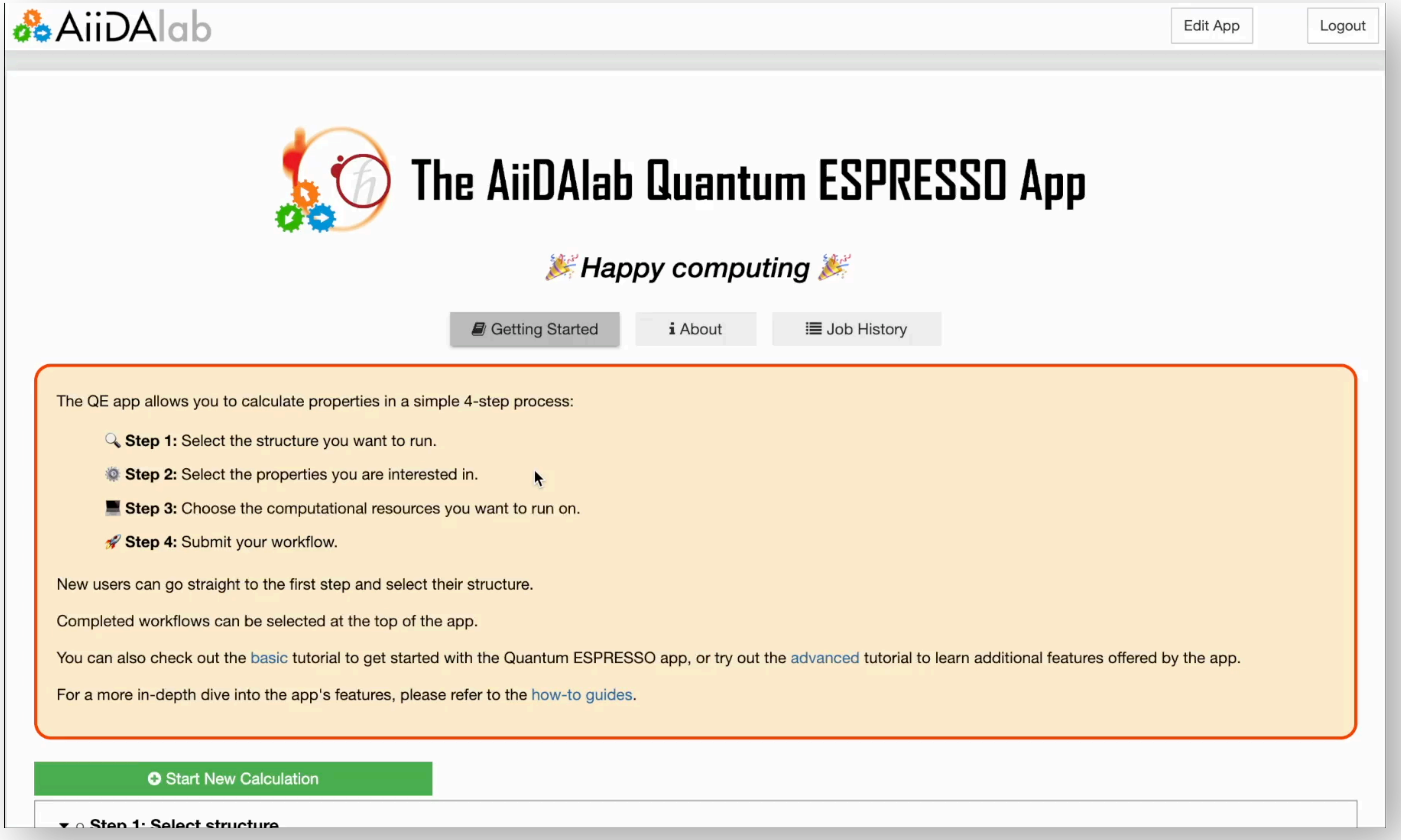


- **Jupyter(Hub)-based**, with AiiDA pre-configured
- **AppMode** to hide input cells, only show outputs as "web apps"
- Installable "Apps" with 1 click from an App Store
- Provide access to custom GUIs, making robust workflows accessible
- One example: **Quantum ESPRESSO app**

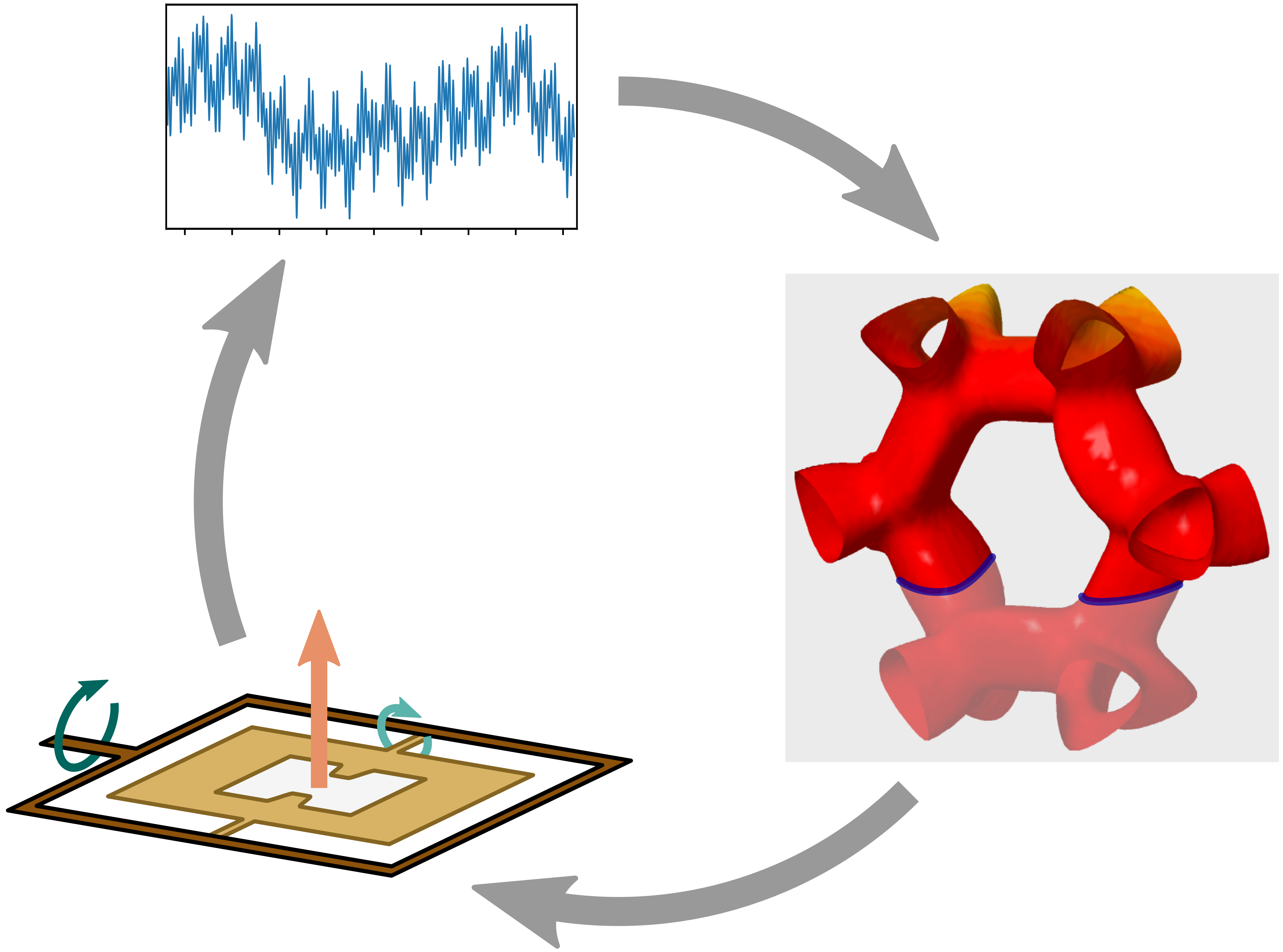
AiiDAlab Quantum ESPRESSO app

Choice of workflows/plugins driven by: existence of robust workflows, broad interest of community (PSI, Empa, ...), ...

https://www.aiidalab.net



Towards (FAIR) autonomous laboratories

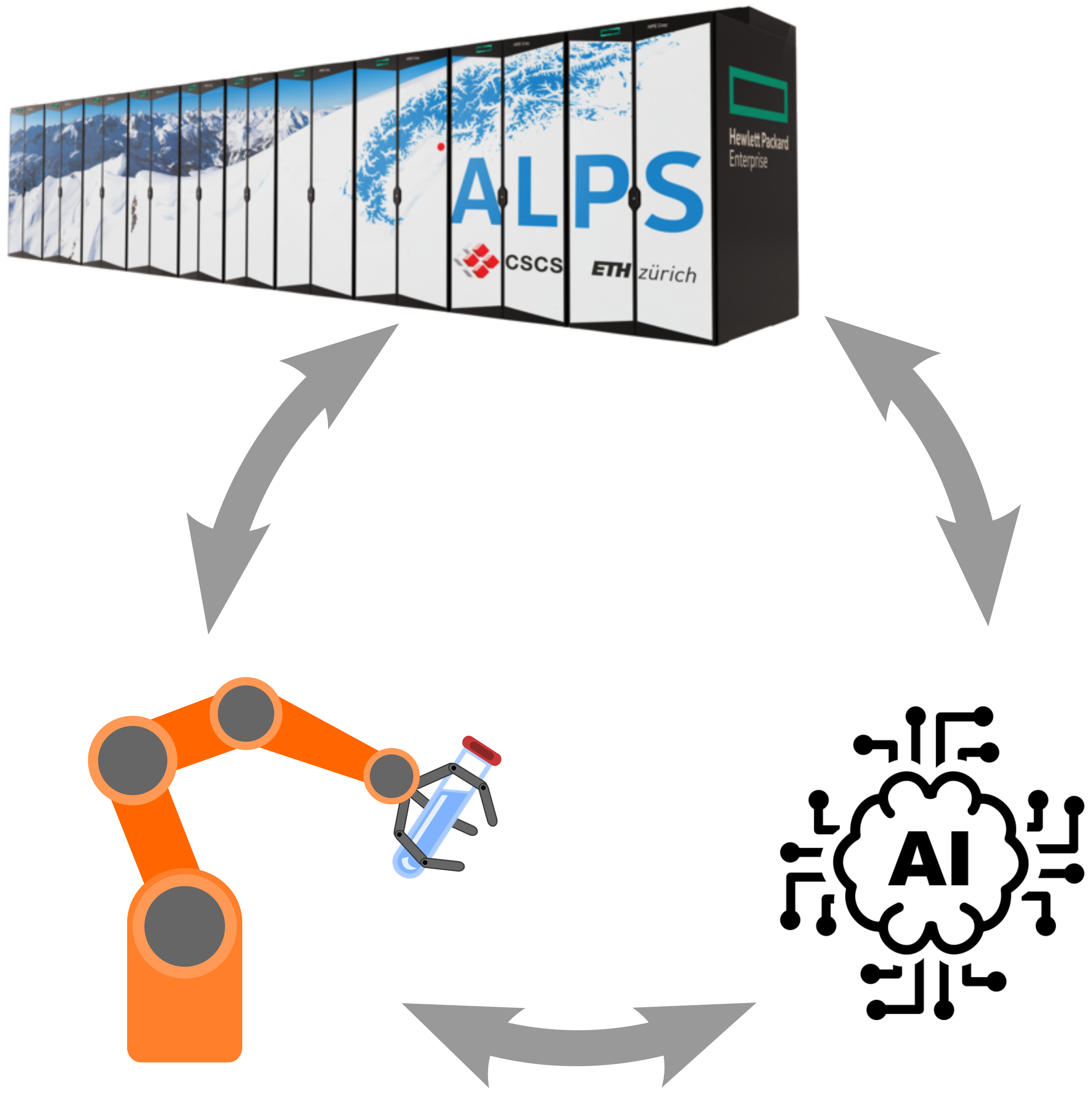


Enabling “FAIR-by-design” autonomous laboratories

PSI

The future of computational materials science:

Automated simulations
via robust workflows



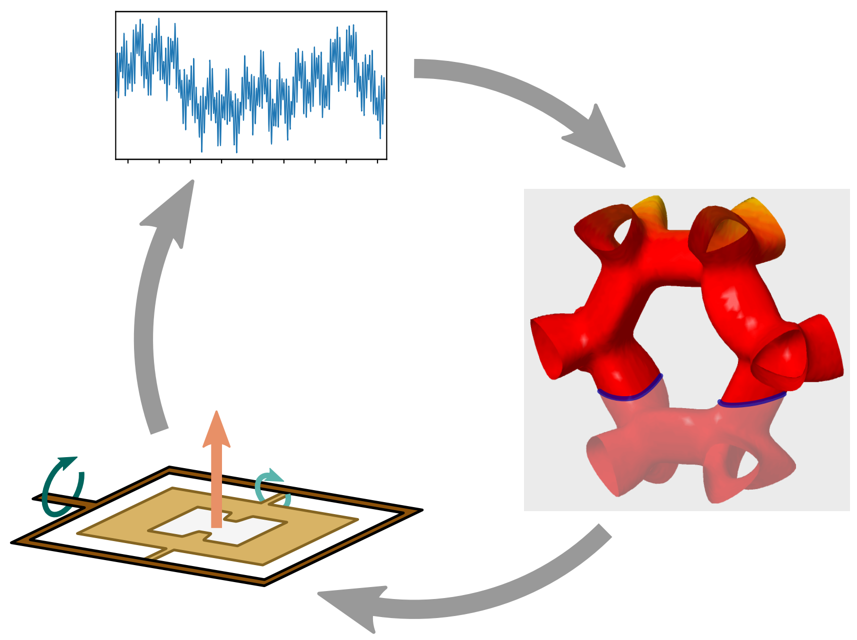
Robotic
experiments

Artificial intelligence/ML
for autonomous decisions

Goal: **open-science platform for accelerated materials research**
enabling **seamless FAIR data collection/sharing “by design”**

"Discover" with autonomous workflows

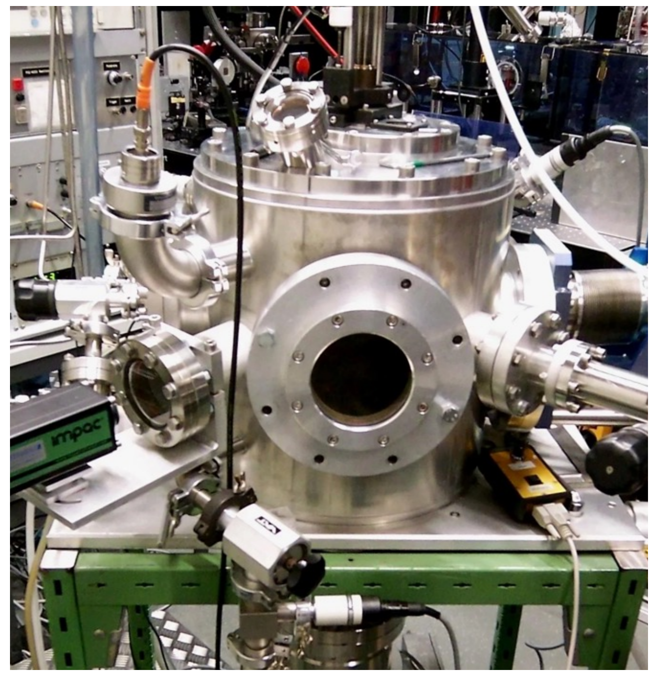
Ongoing projects:



Accurate Fermi-surface mapping (Shubnikov-de Haas)

Collaboration with Philip Moll (MPG, Germany)

Goal: accelerate experiments integrating with robotic 2-axis SdH equipment



Pulsed laser deposition (autonomous crystal growth)

Collaboration with Nikita Shepelin (PSI CNM)

Goal: accelerate identification of ideal parameters for crystalline growth

Battery assembly and testing

Collaboration within BIG-MAP/Battery2030+

M. Vogler et al. **Matter 6**, 2647(2023); **ChemRxiv** chemrxiv-2024-vfq1n (2024)

Collaboration with Corsin Battaglia (Empa)

P. Kraus, .., GP, **J. Mater. Chem. A 12**, 10773 (2024)

Goals: 1. accelerate optimization of batteries (e.g. end-of-life)

2. Develop platform for autonomous orchestration and data management

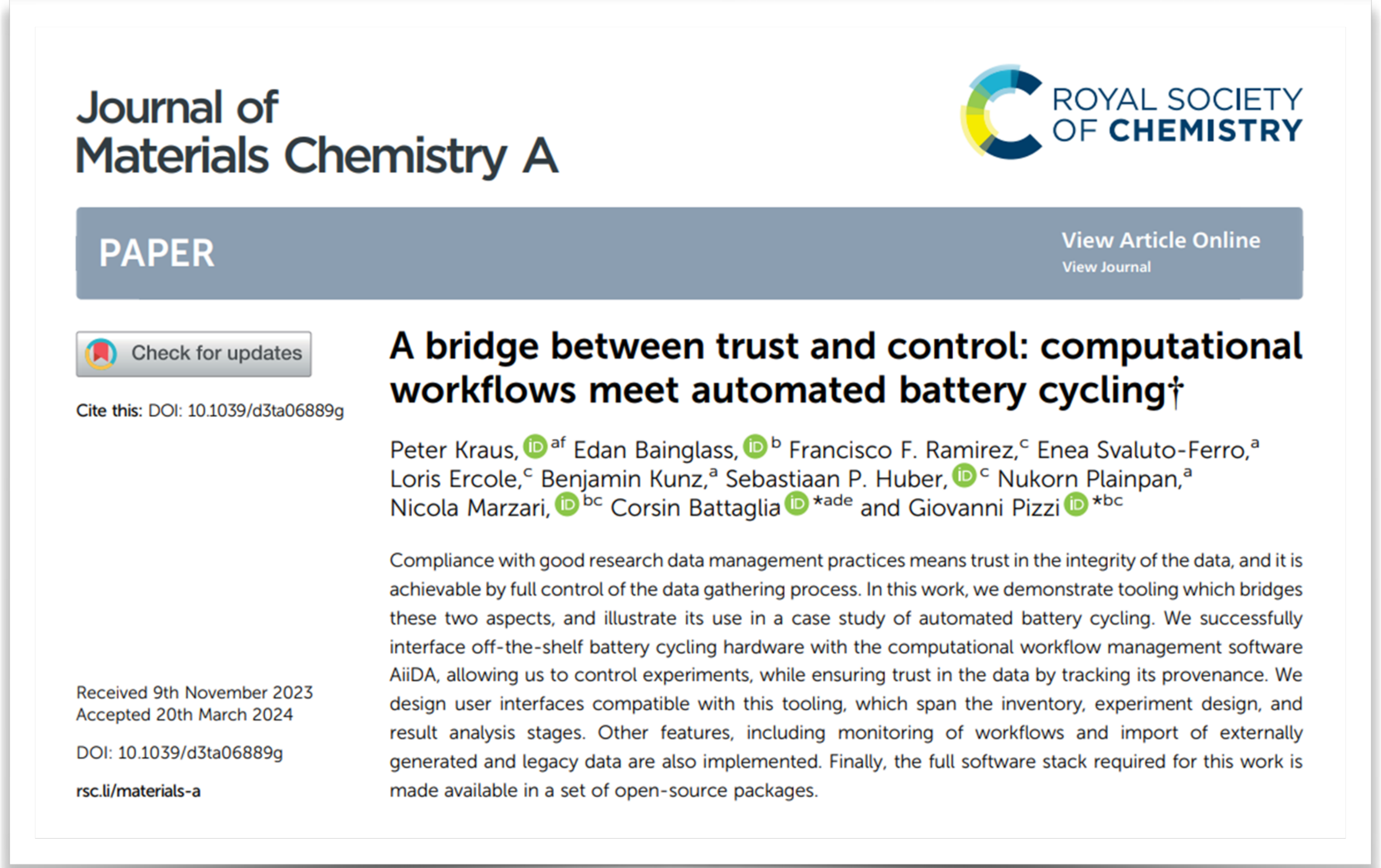
From experiment orchestration to FAIR digital twins

PSI

- **AiiDAlab-Aurora integration**
  - Collaboration with Battaglia's lab (Empa)
  - Automated battery experiments, orchestrated via *AiiDA* (+ *tomato* [1])
  - *AiiDAlab GUI*: batch experiment submission + data analysis

PREMISE

BIG-MAP


Journal of Materials Chemistry A

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Check for updates

Cite this: DOI: 10.1039/d3ta06889g

A bridge between trust and control: computational workflows meet automated battery cycling†

Peter Kraus,[af] Edan Bainglass,[b] Francisco F. Ramirez,[c] Enea Svaluto-Ferro,[a] Loris Ercole,[c] Benjamin Kunz,[a] Sebastiaan P. Huber,[c] Nukorn Plainpan,[a] Nicola Marzari,[bc] Corsin Battaglia[*ade] and Giovanni Pizzi[*bc]

Compliance with good research data management practices means trust in the integrity of the data, and it is achievable by full control of the data gathering process. In this work, we demonstrate tooling which bridges these two aspects, and illustrate its use in a case study of automated battery cycling. We successfully interface off-the-shelf battery cycling hardware with the computational workflow management software AiiDA, allowing us to control experiments, while ensuring trust in the data by tracking its provenance. We design user interfaces compatible with this tooling, which span the inventory, experiment design, and result analysis stages. Other features, including monitoring of workflows and import of externally generated and legacy data are also implemented. Finally, the full software stack required for this work is made available in a set of open-source packages.

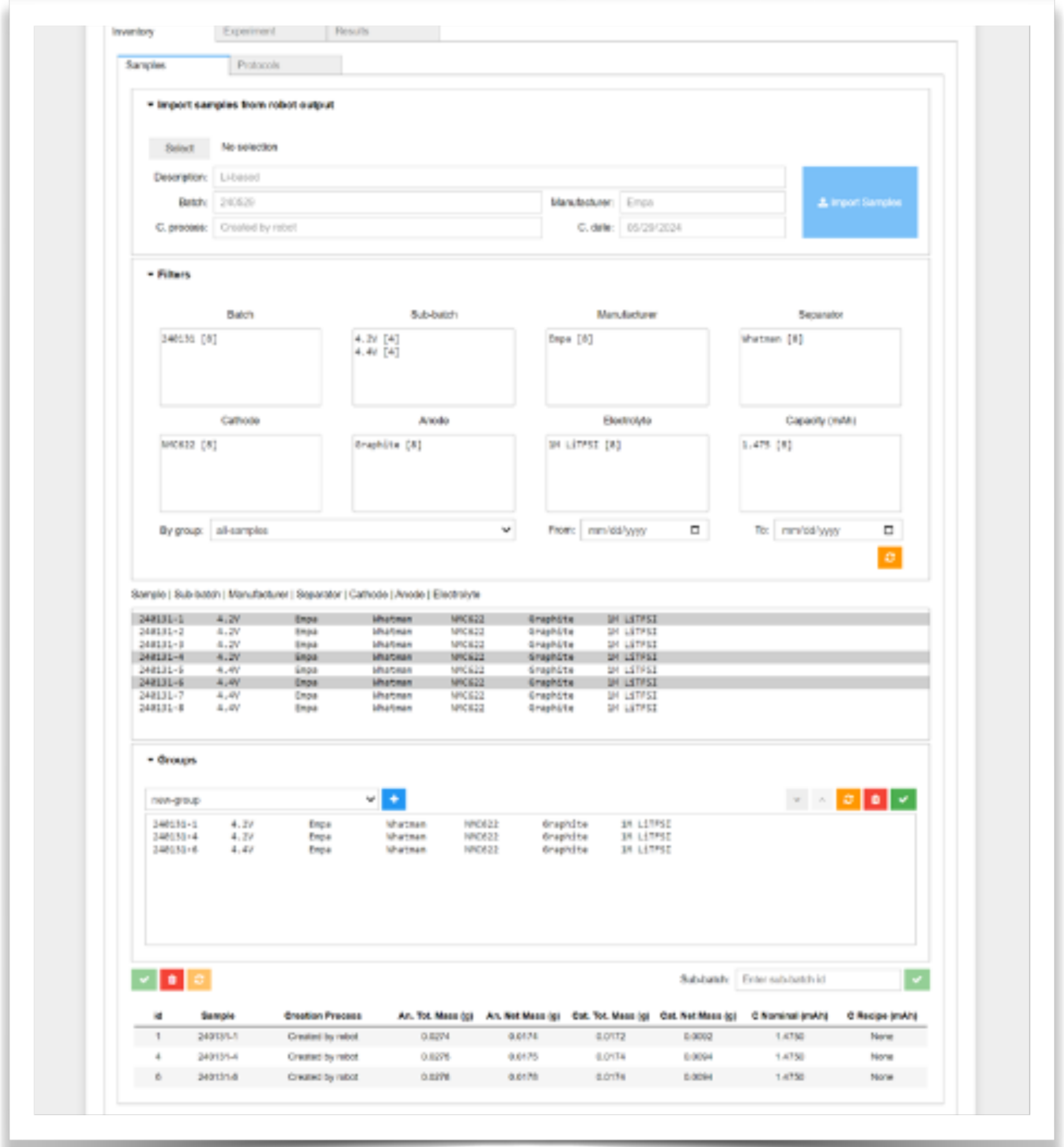
Received 9th November 2023
Accepted 20th March 2024

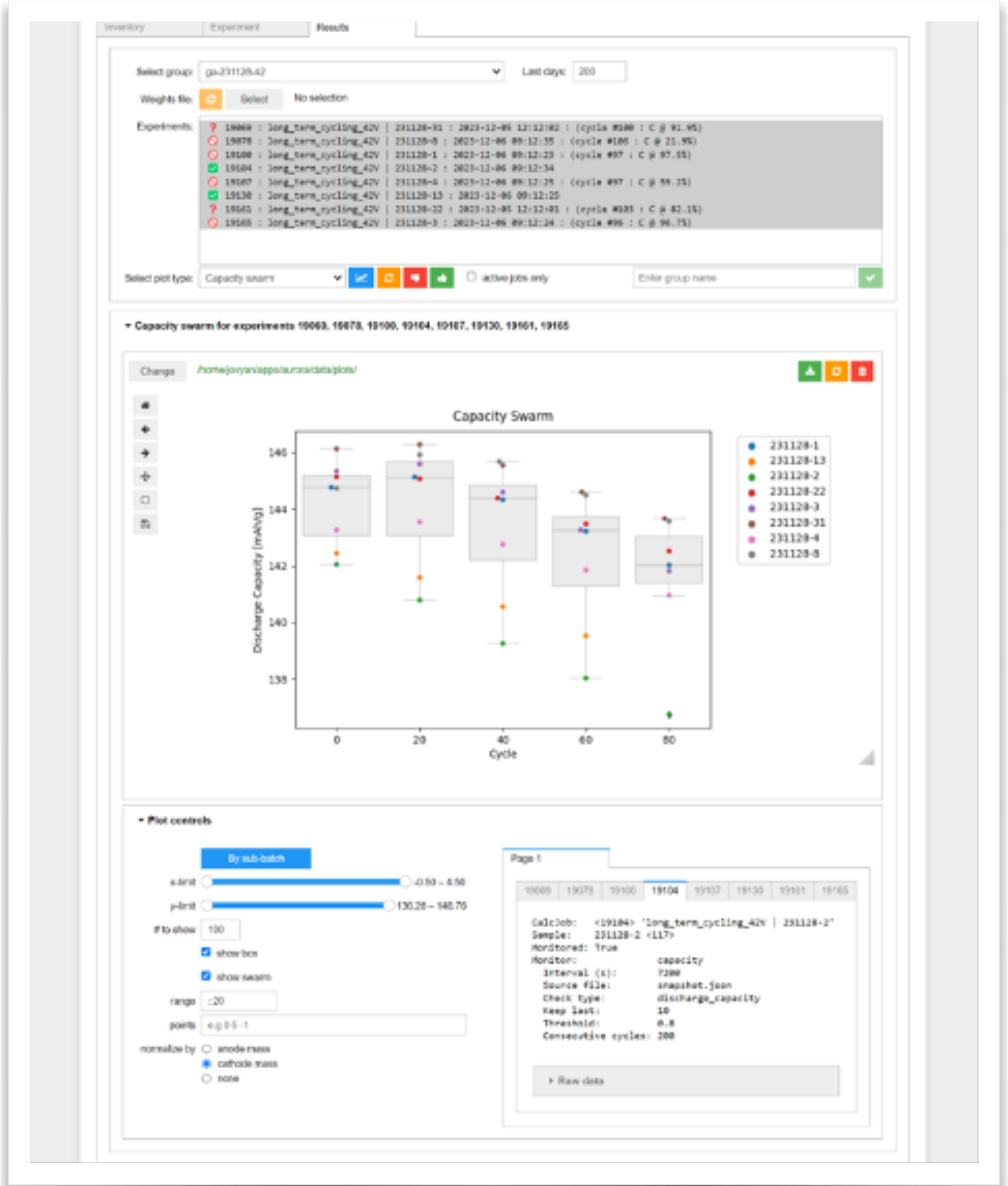
DOI: 10.1039/d3ta06889g

rsc.li/materials-a

P. Kraus *et al.*, **J. Mater. Chem. A 12**, 10773 (2024)

(co-funded by BIG-MAP)





AiiDAlab-Aurora GUI

(video tutorials: https://ord-premise.org/tutorials/)

[1] https://github.com/dgbowl/tomato

From experiment orchestration to FAIR digital twins

PSI

- **AiiDAl**
  - Colla
  - Auto orch
  - *AiiDA* subn

SE

BIG-MAP

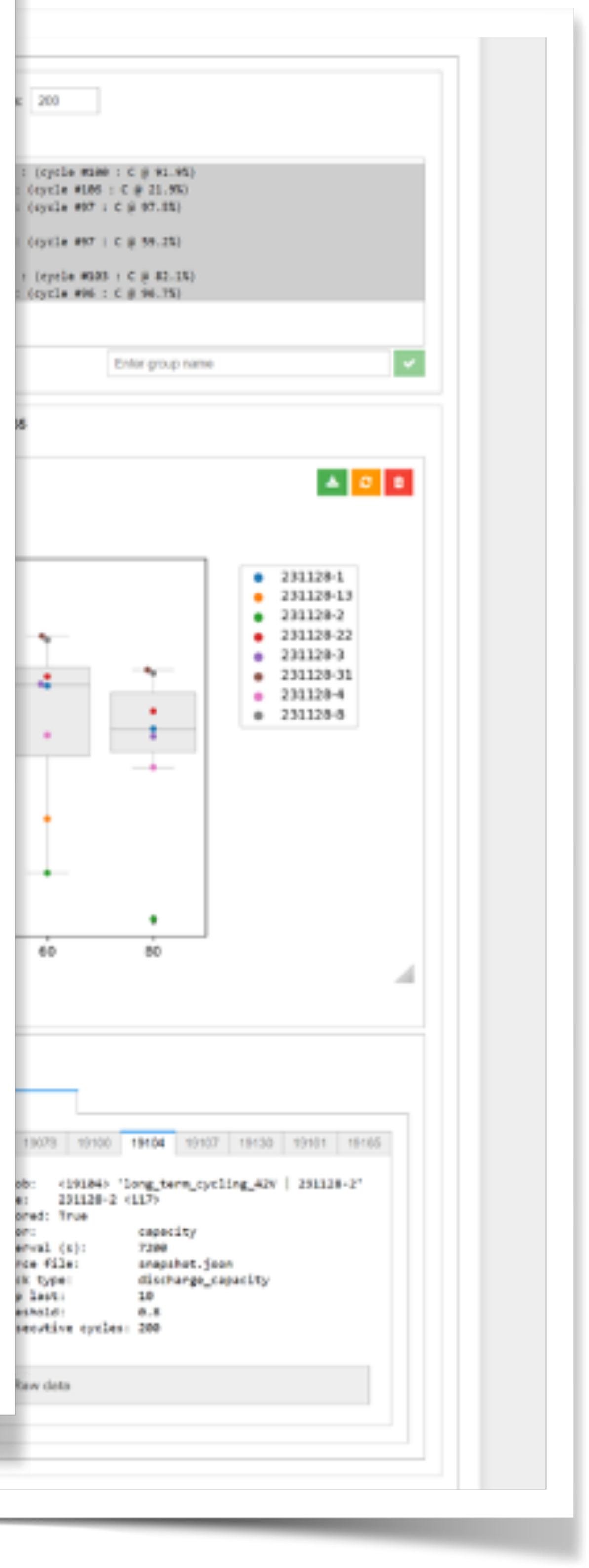
Next step: integration with Bayesian algorithms for autonomous optimization of battery electrolyte formulation

Ultimate goal (Research): accelerate design of improved batteries

Ultimate goal (Platform): Seamless data interoperability (experiments and simulations) and autonomous platform "FAIR-by-design"

See more on the PREMISE website
https://ord-premise.org/deliverables/

PREMISE



P. Kraus *et al.*, **J. Mater. Chem. A 12**, 10773 (2024)
(co-funded by BIG-MAP)

AiiDAlab-Aurora GUI
(video tutorials: https://ord-premise.org/tutorials/)

[1] https://github.com/dgbowl/tomato

FAIRness in practice? Considerations from my own experience

PSI

- Many (but not all!) researchers **care about maximizing output** (papers/citations/faster tasks)
- **Tool availability** (ELNs, workflow managers, …) (if well designed!) **can boost FAIR adoption**
- Design **tools around researchers' needs**:
  - Not all researchers are professional coders: **make it easy, clarify immediate benefits**
  - Clarify **target audience** (to you and stakeholders)! Limit use cases to avoid "failure"
    E.g. *simulation experts vs experimentalists;* but also target technical goals:
    performance vs. easy-of-use vs. reproducibility vs. dynamic workflows vs. …
  - If you have a team: balance RSEs with background/**experience** in the field (important!) with **professional developers**, and encourage mutual learning
  - **RSE role** still far from being consolidated: tension between "doing research" and robust software development (also for funding)

Acknowledgements: funding

PSI

MARVEL
NATIONAL CENTRE OF COMPETENCE IN RESEARCH

Swiss National Science Foundation

SNSF NCCR "MARVEL"
Discovery of new materials via simulations and dissemination of curated data

MAX DRIVING THE EXASCALE TRANSITION

H2020 Centre of Excellence "MaX"
Scaling towards exascale machines and high-throughput efficiency

Moreover:

swissuniversities

ETH BOARD

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

State Secretariat for Education, Research and Innovation SERI

BIG-MAP

MarketPlace

nffa.eu
nanoscience foundries & fine analysis

DOME 4.0

INTERSECT

EMMC

Open Model

Acknowledgements: AiiDA, AiiDAlab, Materials Cloud teams

PSI

The current AiiDA, AiiDAlab and Materials Cloud teams

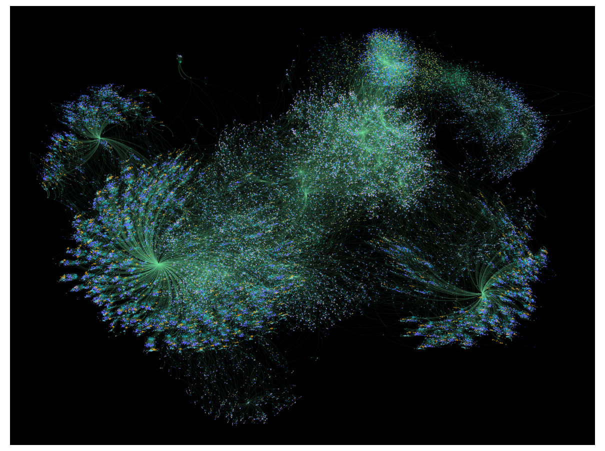
Edan Bainglass (PSI) | Roberto Bendinelli (EPFL) | Marnik Bercx (EPFL) | Miki Bonacci (PSI) | Dou Du (EPFL) | Kristjan Eimre (EPFL) | Guillaume Fraux (EPFL) | Julian Geiger (PSI) | Alex Goscinski (PSI) | Valeria Granata (EPFL) | Daniel Hollas (U. Bristol) | Sebastiaan P. Huber (Microsoft) | Ali Khosravi (PSI)

François Liot (EPFL) | Nicola Marzari (EPFL, PSI) | Carlo Pignedoli (Empa) | Giovanni Pizzi (PSI) | Louis Ponet (EPFL) | Francisco F. Ramirez (EPFL) | Thomas Schulthess (ETHZ,CSCS) | Chris Sewell (EPFL) | Leopold Talirz (Microsoft) | Joost VandeVondele (ETHZ,CSCS) | Xing Wang (PSI) | Aliaksandr Yakutovich (Empa) | Jusong Yu (PSI)

And all former team members, who actively contributed with ideas and code to make these platforms what they are today:

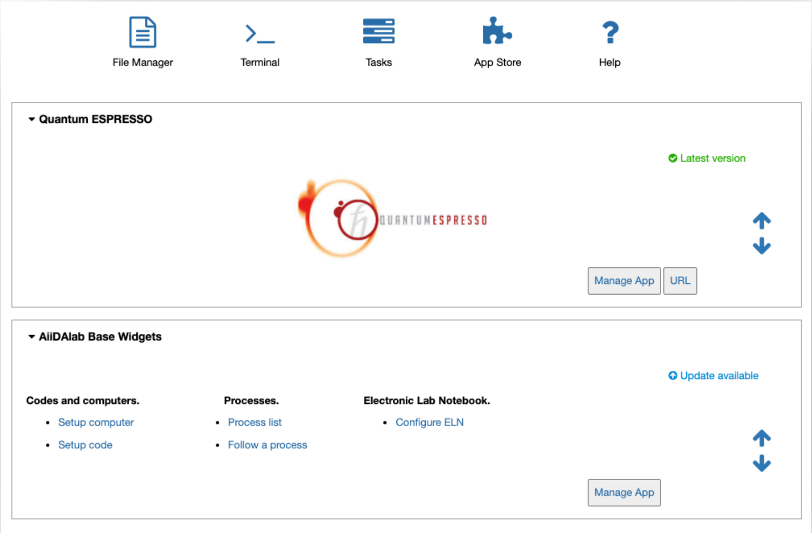
- Carl S. Adorf
- Casper W. Andersen
- Marco Borelli
- Andrea Cepellotti
- Edward Ditler
- Fernando Gargiulo
- Dominik Gresch
- Rico Häuselmann
- Conrad Johnston
- Leonid Kahle
- Snehal Kumbhar
- Boris Kozinsky
- Andrius Merkys
- Nicolas Mounet
- Tiziano Müller
- Elsa Passaro
- Daniele Passerone
- Riccardo Sabatini
- Ole Schütt
- Berend Smit
- Martin Uhrin
- Spyros Zoupanos
- ...

Summary

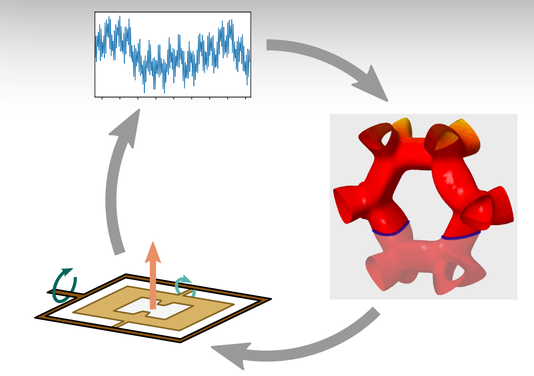


- AiiDA+Materials Cloud: automated simulations and FAIR access to simulation data

- Plugin ecosystem + common workflow interfaces for code interoperability



- Robustness is crucial for users (not only performance)
- Beyond FAIR data to FAIR workflows: robust turn-key workflows, made accessible via AiiDAlab GUIs



- Ontologies can enable FAIR-by-design autonomous labs

AiiDA https://www.aiida.net

MATERIALSCLOUD https://www.materialscloud.org

AiiDAlab https://www.aiidalab.net