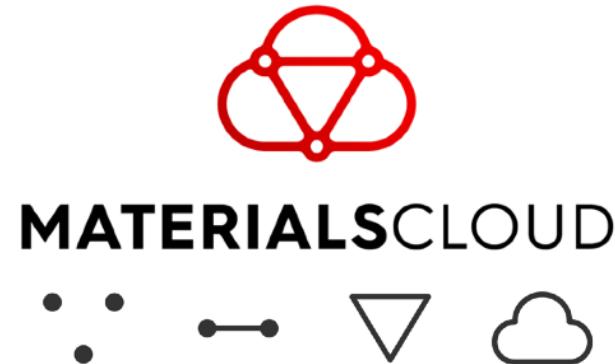
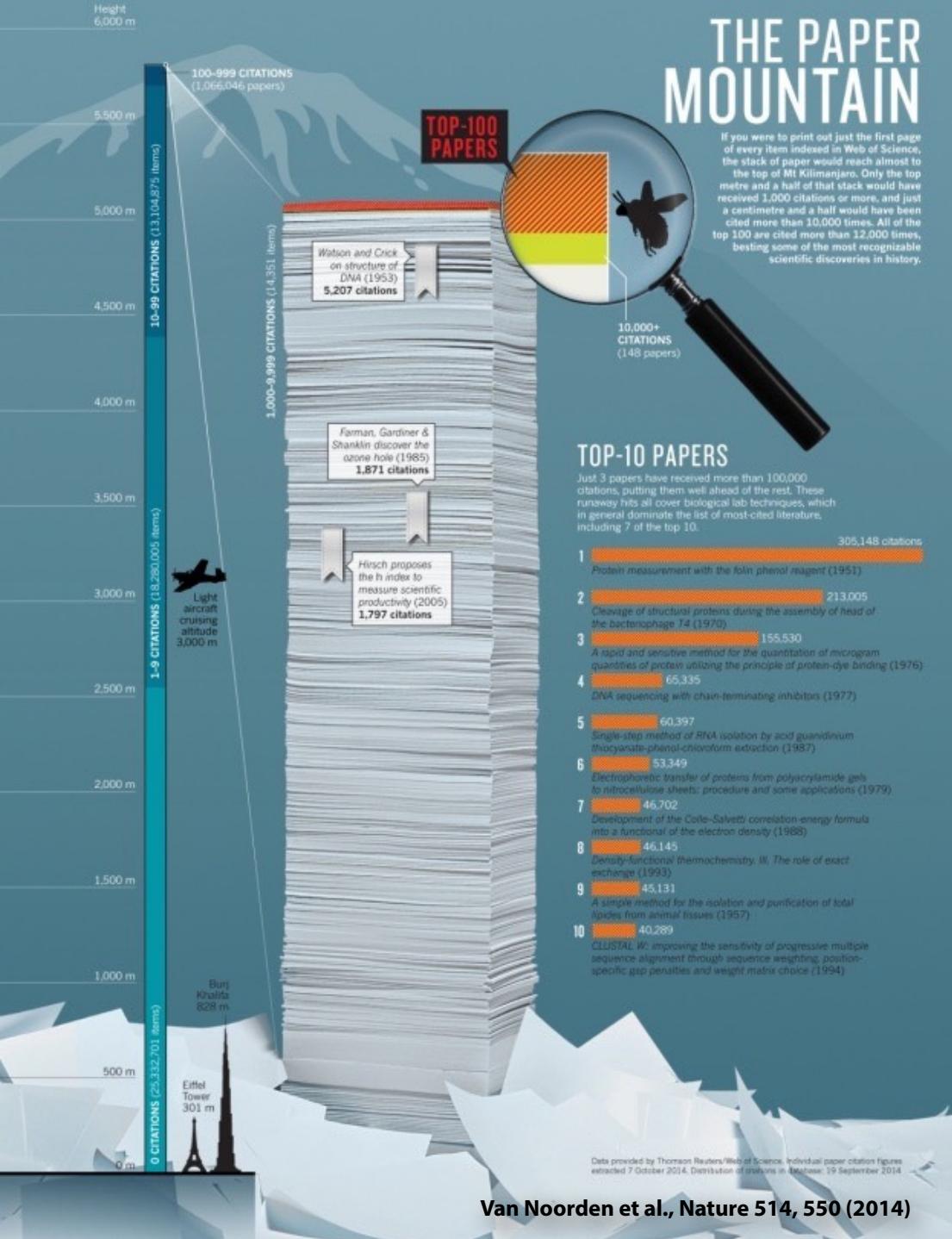


Open Science Platform for Materials Science: AiiDA and the Materials Cloud

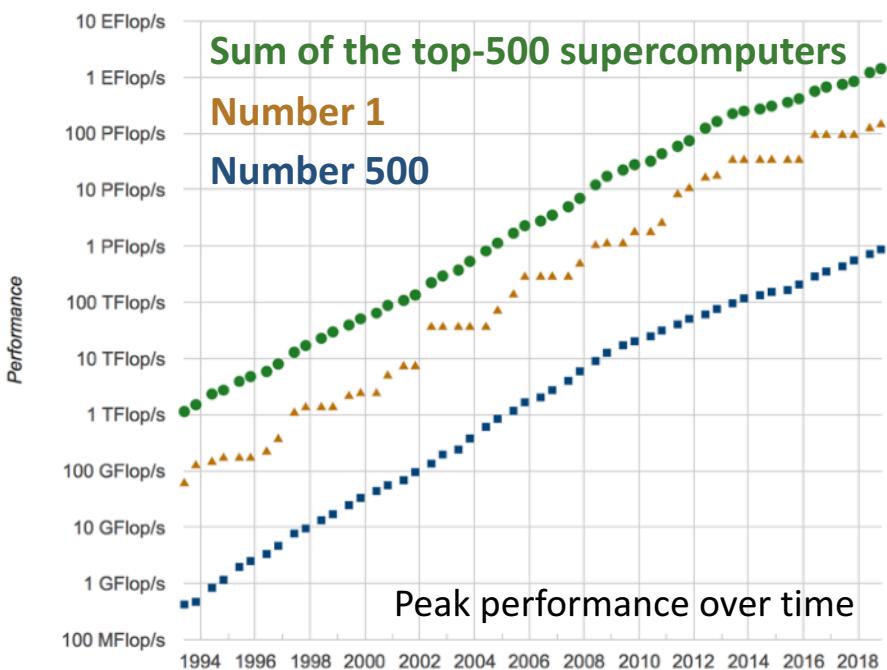
Giovanni Pizzi (EPFL)





Our research: Materials simulations

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



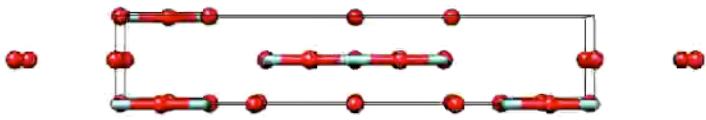
Accuracy and
predictive power
of quantum
engines

**150,000x increase
in the past 20 years**

1 month (1998)
↓
10 seconds (2018)

*Result: materials design and discovery via
high-throughput computations*

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**



Open Science Platform: definition

- Our definition of an Open Science Platform [1]:
 - **Open simulation codes**
 - **Open architecture** to manage simulations and **open workflows**
 - Support for **Open Data, Data Management Plans** and FAIR-compliant sharing
 - **Straightforward availability** of the tools, with **curated open-data services** enabling turn-key workflows (pseudopotential libraries, ...)

[1] Pizzi G. (2018) *Open-Science Platform for Computational Materials Science: AiiDA and the Materials Cloud*. In: Andreoni W., Yip S. (eds), *Handbook of Materials Modeling* (Springer, Cham).



Our goal

Build an open-science infrastructure
with computational services offered
to scientific, industrial community and beyond

Like a synchrotron, but for
open and reproducible simulations



Our two core infrastructures

AiiDA as the “operating system” to manage, automate and store simulations and their results

and

Materials Cloud as the open-science dissemination portal and cloud simulation platform



OPEN SCIENCE PLATFORM:



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



AiiDA and the Materials Cloud

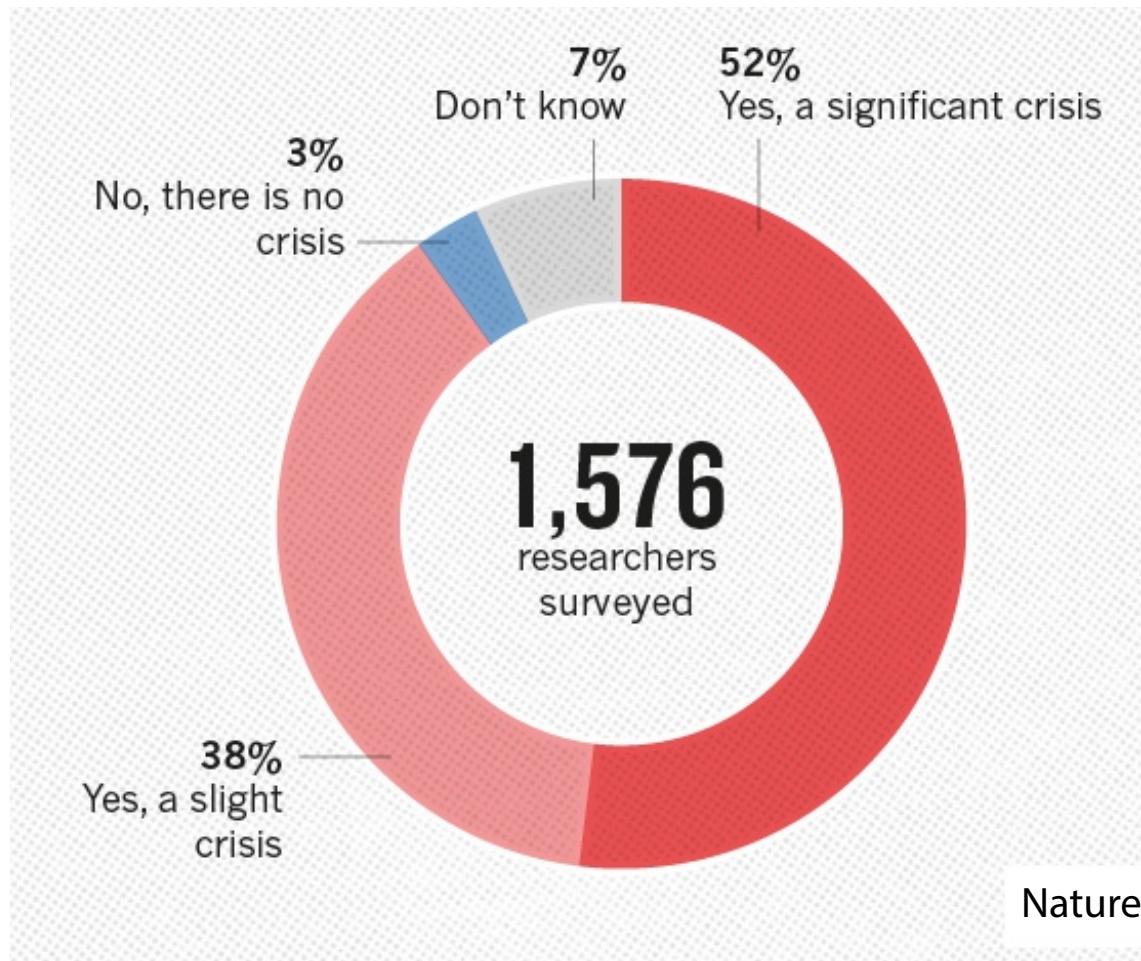
How to manage simulations and their provenance?



We need a **tool** to help us
organise research and store provenance

Reproducibility: a cornerstone of the scientific method

IS THERE A REPRODUCIBILITY CRISIS?



Nature 533, 452–454 (2016)



AiiDA and the Materials Cloud

Reproducibility: a cornerstone of the scientific method

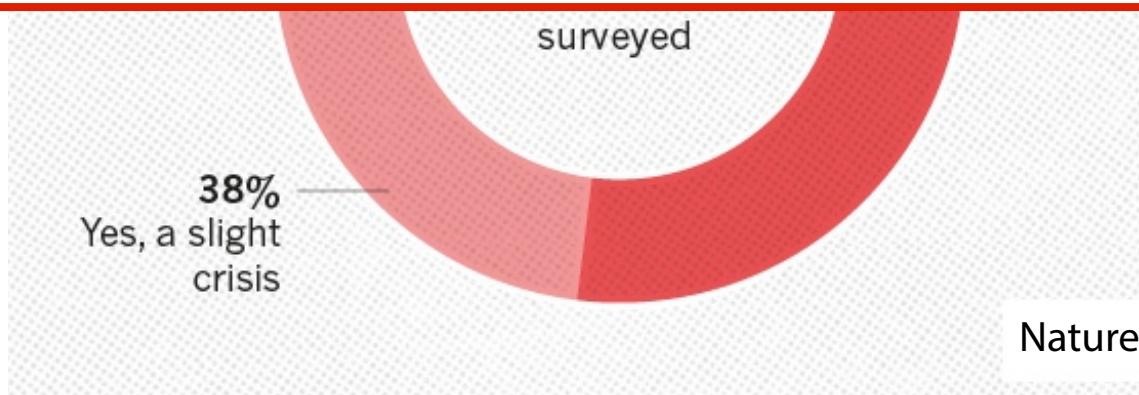
IS THERE A REPRODUCIBILITY CRISIS?



No excuses in computational science

We can and **must** be fully reproducible

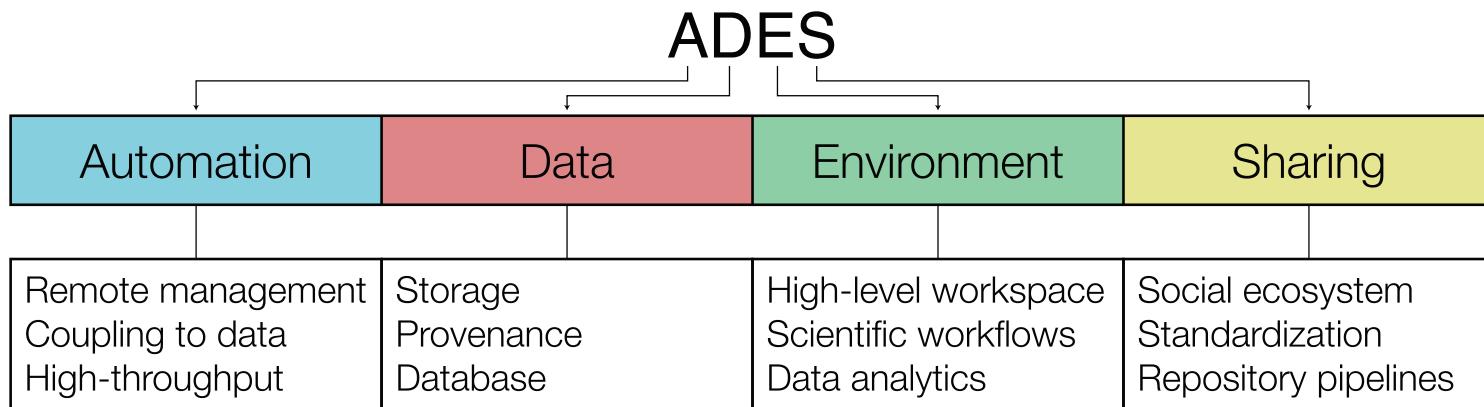
CHALLENGE #2: make open-science **easier**



Nature **533**, 452–454 (2016)



AiiDA development timeline



This repository Search Pull requests Issues Marketplace Explore

alidateam / aiida_core Unwatch 14 Star 61 Fork 46

Code Issues 154 Pull requests 3 Projects 2 Wiki Insights Settings

The official repository for the AiiDA code Edit

8,751 commits 3 branches 31 releases 37 contributors

Branch: develop New pull request Create new file Upload files Find file Clone or download

sphuber Merge pull request #1478 from sphuber/fly_1477_process_on_except_form... Latest commit 543463c 17 hours ago

jenkins-data Merge branch 'develop' into jenkins-workflows a day ago

.travis-data Fixing the coverage report for the two backends (independently) 4 days ago

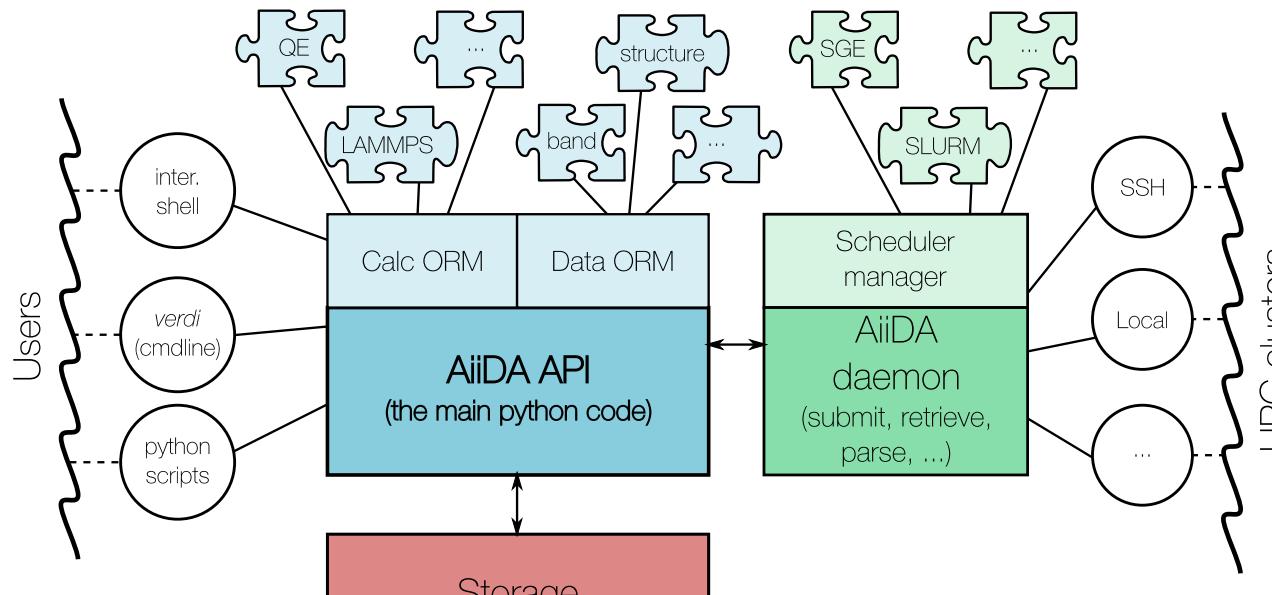
G. Pizzi et al., Comp. Mat. Sci. 111, 218-230 (2016)

<http://www.aiida.net>



AiiDA and the Materials Cloud

AiiDA



Main features

- Python (2.7 & 3.6) infrastructure
- SQL database backend, access via a Python ORM
- Local connection to clusters, or via ssh using a python API
- Interface to various job schedulers (SGE, Torque, LSF, PBS Pro, SLURM, ...)
- Event-based daemon with remote management and workflow execution manager
- REST APIs using Flask to expose one own's data
- Plugin management system and extended code support
- Easy sharing of the results with other users in the community



AiiDA



AiiDA and the Materials Cloud



AiiDA submission

```
code = Code.get('pw-6.3@daint-mr25')
builder = code.new_builder()

builder.metadata.options = {
    'max_wallclock_seconds': 600,
    'max_wallclock_seconds': {"num_machines": 2}}

Structure = DataFactory('structure')
structure = Structure(ase = read('TiO2.cif'))

Dict = DataFactory('dict')
parameters = Dict(dict={
    'CONTROL': {
        'calculation': 'scf',
        'restart_mode': 'from_scratch'},
    'SYSTEM': {'ecutwfc': 40.}})

Kpoints = DataFactory('array.kpoints')
kpoints = Kpoints(kpoints_mesh = [4,4,4])

builder.structure = structure
builder.parameters = parameters
builder.kpoints = kpoints
builder.pseudos = get_pseudos_from_family(structure,
    'SSSP_efficiency_v1.0')

aiida.engine.submit(builder)
```



AiiDA submission

```
code = Code.get('pw-6.3@daint-mr25')
builder = code.new_builder()

builder.metadata.options = {
    'max_wallclock_seconds': 600,
    'max_wallclock_seconds': {"num_machines": 2} }

Structure = DataFactory('structure')
structure = Structure(ase = read('TiO2.cif'))

Dict = DataFactory('dict')
parameters = Dict(dict={
    'CONTROL': {
        'calculation': 'scf',
        'restart_mode': 'from_scratch'},
    'SYSTEM': {'ecutwfc': 40.}})

Kpoints = DataFactory('array.kpoints')
kpoints = Kpoints(kpoints_mesh = [4,4,4])

builder.structure = structure
builder.parameters = parameters
builder.kpoints = kpoints
builder.pseudos = get_pseudos_from_family(structure,
    'SSSP_efficiency_v1.0')

aiida.engine.submit(builder)
```

Switch computers in one line
supports different schedulers,
version of codes, ...

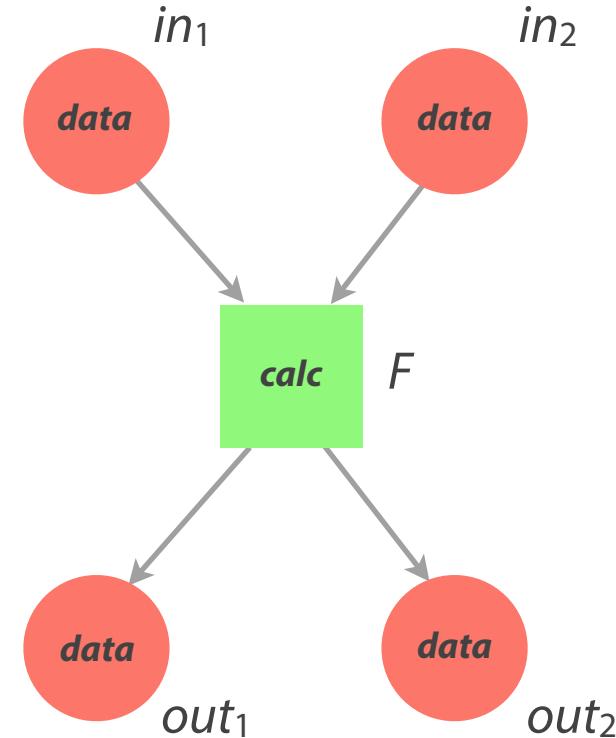
Define (only) necessary inputs
Interface designed by plugin

**Inputs stored in the DB, and
handing over to the daemon**

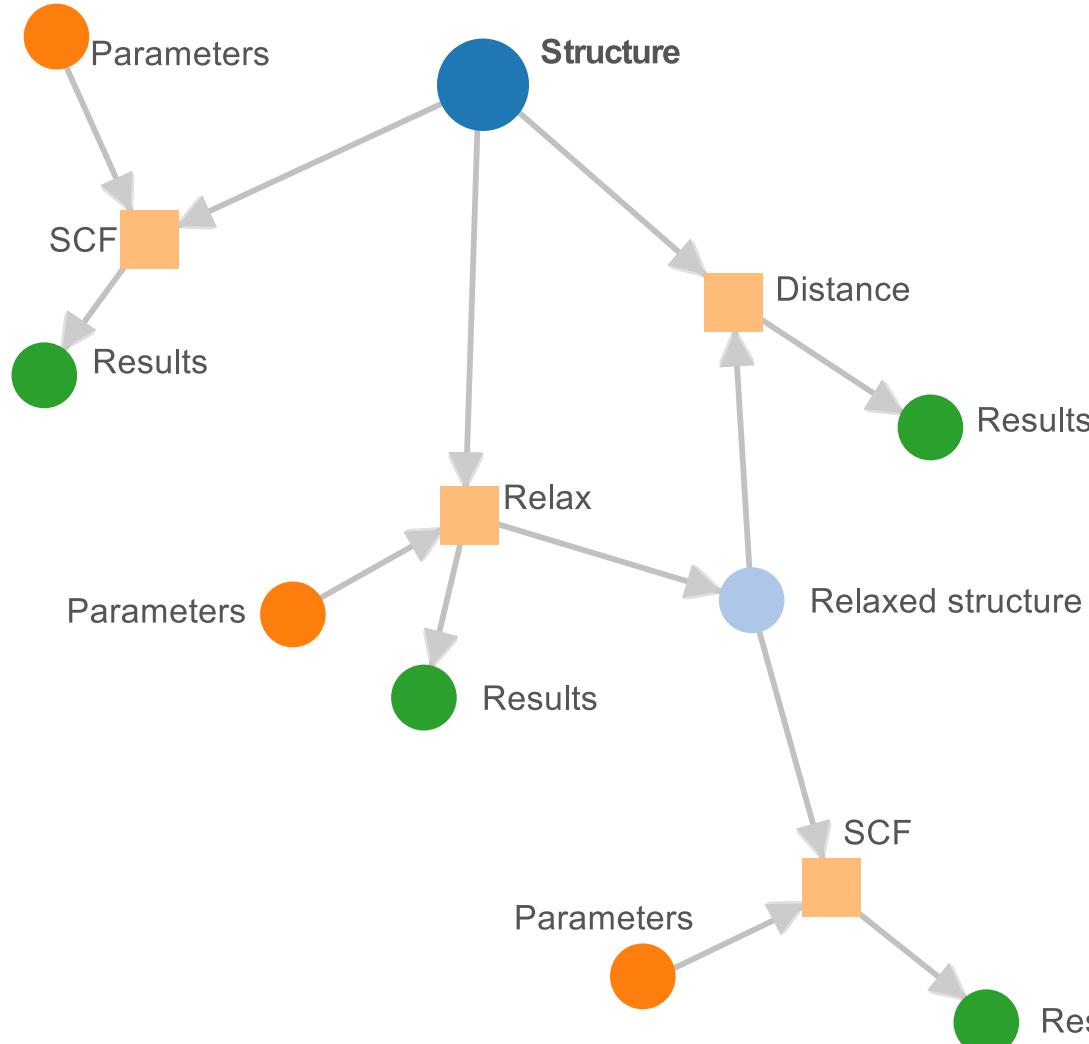


Storage and provenance

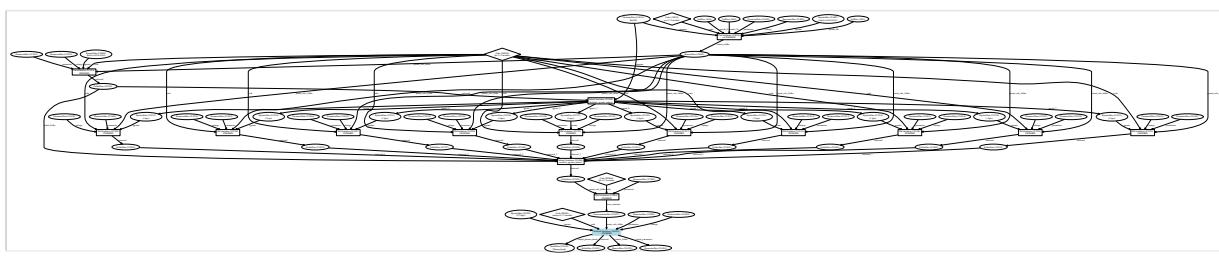
- *Calculated properties*: result of complex, connected calculations
- How do we store simulations **preserving the connected structure?**
- Inspiration from the *open provenance model*
- **Any calculation: a function**,
converting inputs to outputs:
$$\mathbf{out}_1, \mathbf{out}_2 = F(\mathbf{in}_1, \mathbf{in}_2)$$
- **Each object is a node in a graph**,
connected by directional labeled links
- Output nodes can be used as inputs



Data provenance: Directed Acyclic Graphs

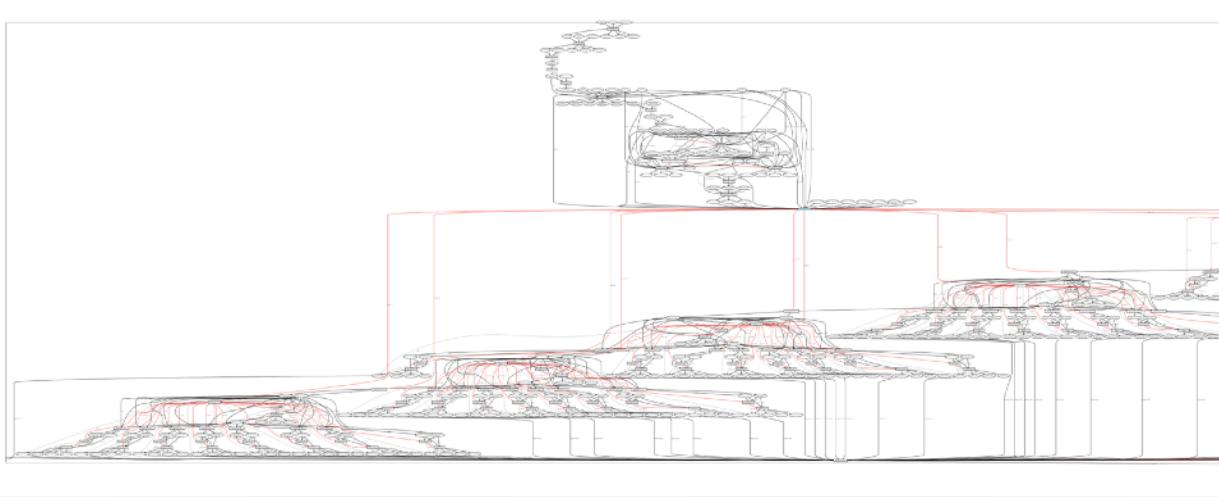


“Simple” graphs of workflows for a single material



Phonon dispersion

(atom oscillations around equilibrium positions:
thermal transport,
electronic mobility, ...)



Molecular dynamics of Lithium in a solid electrolyte

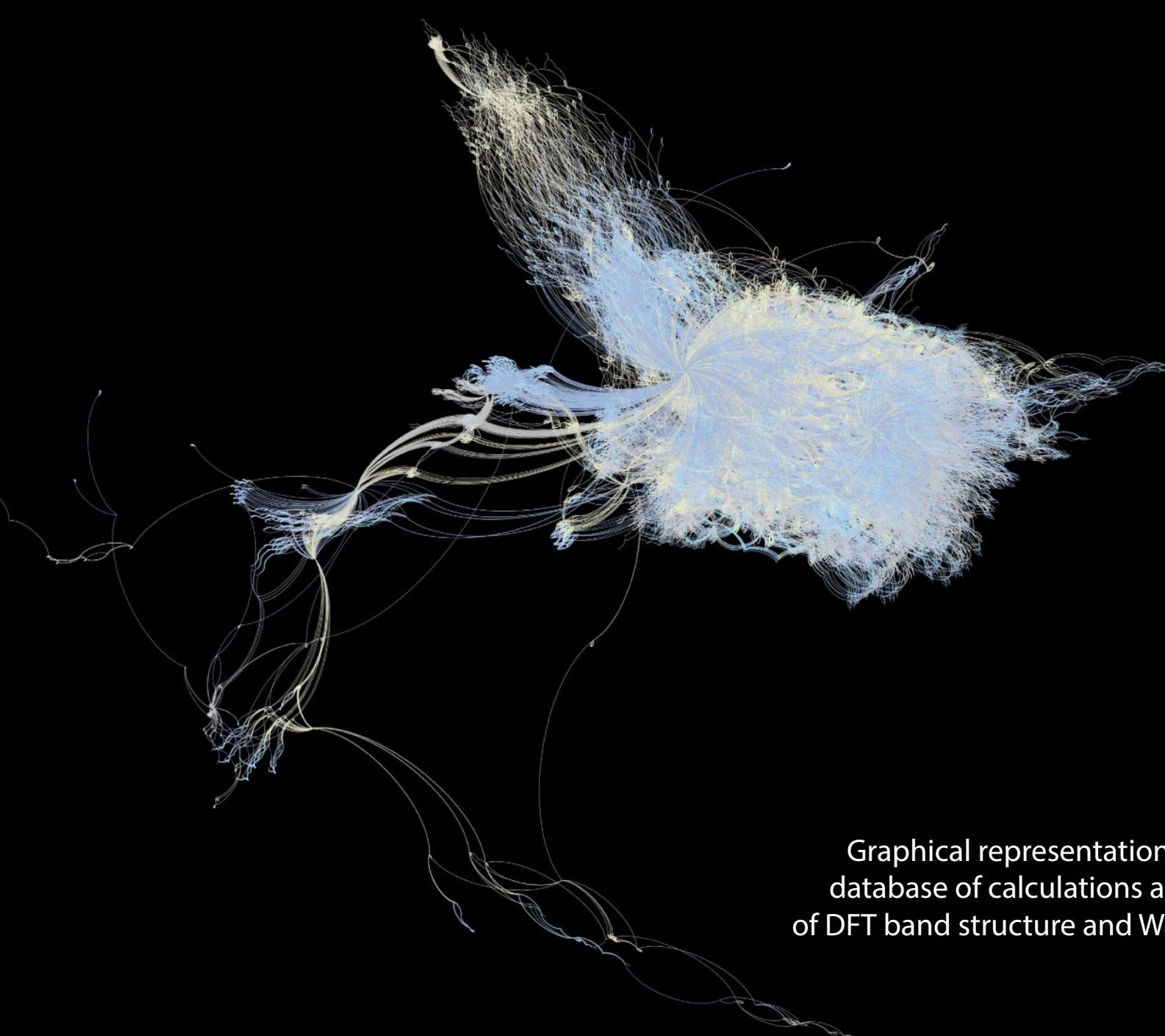
(Discover novel, safe and
efficient electrolytes for Li-
batteries)



Elastic constants

(response of materials to
stresses and deformations)

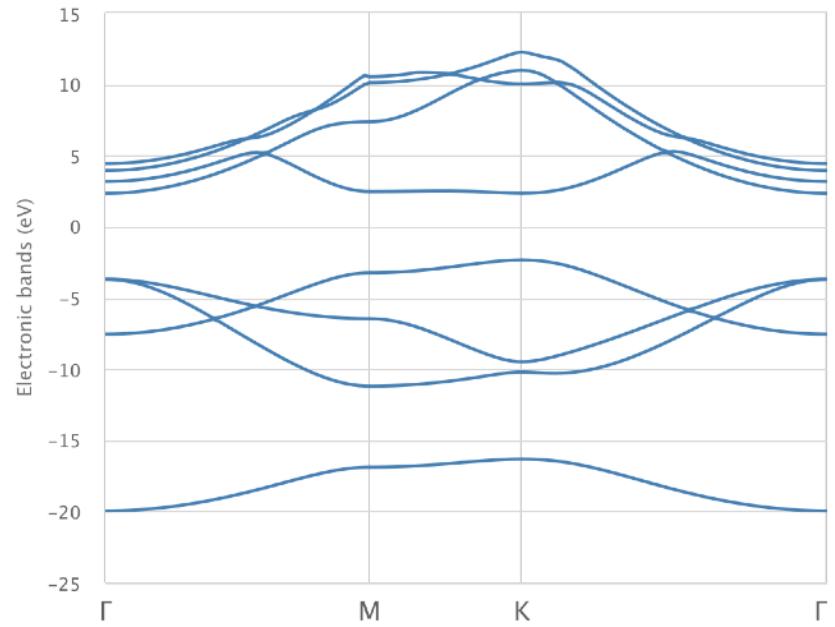
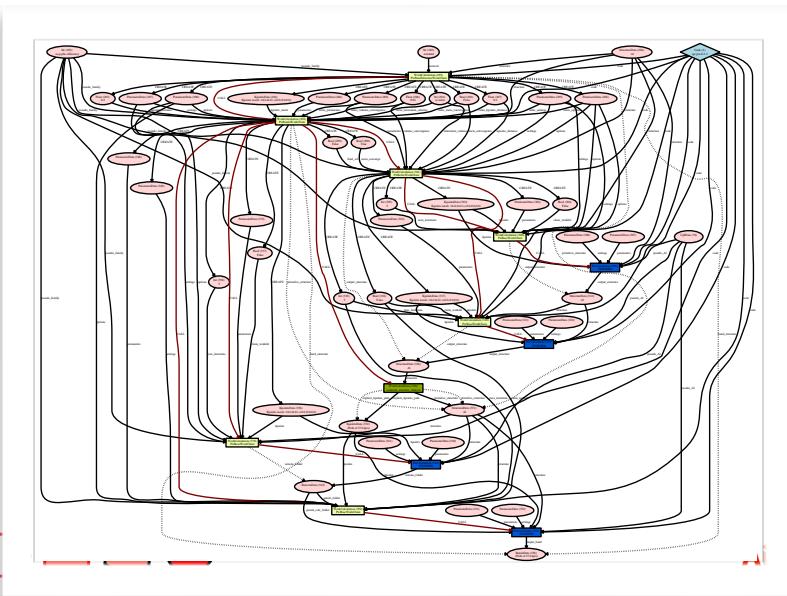
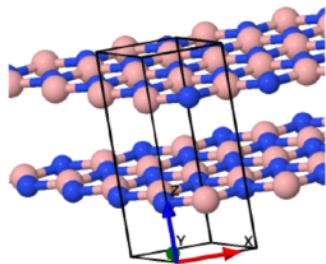




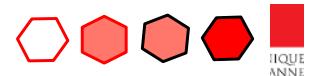
Graphical representation of an AiiDA
database of calculations and workflows
of DFT band structure and Wannier functions

Turn-key workflows in AiiDA

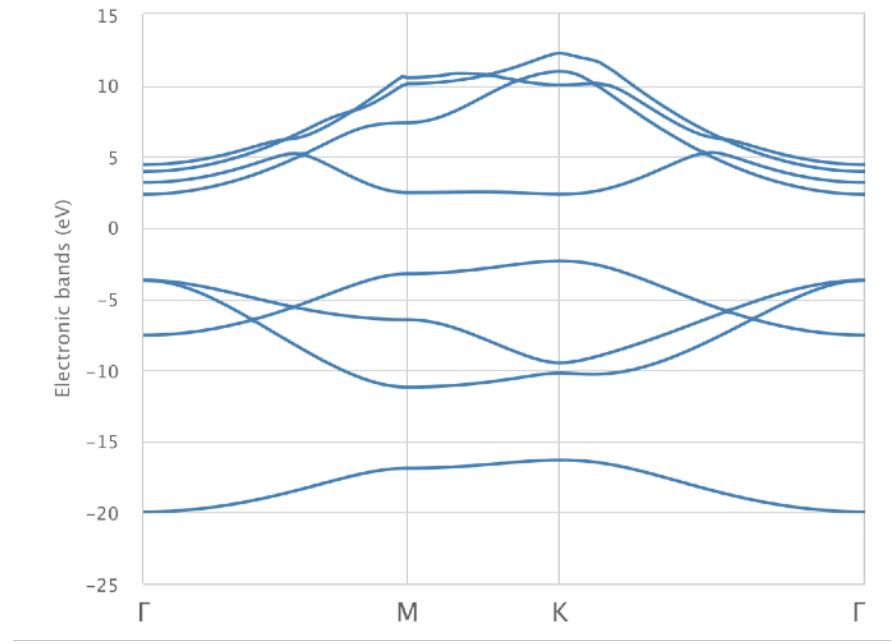
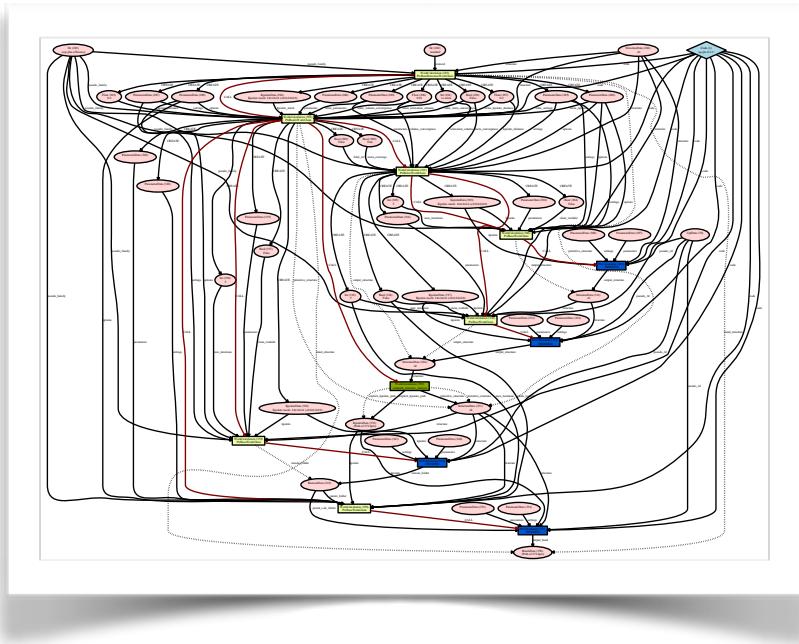
- Given a material, we often need to compute advanced quantities
- These are often non-trivial and result from a complex workflow



AiiDA and the Materials Cloud



Turn-key workflows in AiiDA



- The AiiDA provenance graph allows to know how the structure was computed and to **reproduce that single specific calculation**: *log of “what happened in the past”*
- We need also an **easy way to re-run the same calculation again** with different parameters or for a different material: **turn-key workflows**



Turn-key workflows in AiiDA

```
class PwBandsWorkChain(WorkChain):
    @classmethod
    def define(cls, spec):

        spec.input('code',
                   valid_type=Code)
        spec.input('structure',
                   valid_type=StructureData)
        spec.input('pseudo_family',
                   valid_type=Str)

        spec.outline(
            cls.setup,
            cls.validate_inputs,
            if_(cls.should_do_relax)(
                cls.run_relax,
            ),
            cls.run_seekpath,
            cls.run_scf,
            cls.run_bands,
            cls.results,
        )
```

- “**Operating system**” for all calculations
- Automatic provenance tracking in the DB
- Control provenance granularity store level of detail relevant to the workflows
- Progress checkpointing
restart from arbitrary step, retry on failure, allows to shut down daemon and continue later
- Easy debugging, self-documenting



Turn-key workflows in AiiDA

```
class PwBandsWorkChain(WorkChain):
    @classmethod
    def define(cls, spec):

        spec.input('code',
                   valid_type=Code)
        spec.input('structure',
                   valid_type=StructureData)
        spec.input('pseudo_family',
                   valid_type=Str)

        spec.outline(
            cls.setup,
            cls.validate_inputs,
            if_(cls.should_do_relax)(
                cls.run_relax,
            ),
            cls.run_seekpath,
            cls.run_scf,
            cls.run_bands,
            cls.result
        )
```

- “**Operating system**” for all calculations
- Automatic provenance tracking in the DB
- Control provenance granularity store level of detail relevant to the workflows
- Progress checkpointing restart from arbitrary step, retry on failure, allows to shut down daemon and continue later
- Easy debugging, self-documenting

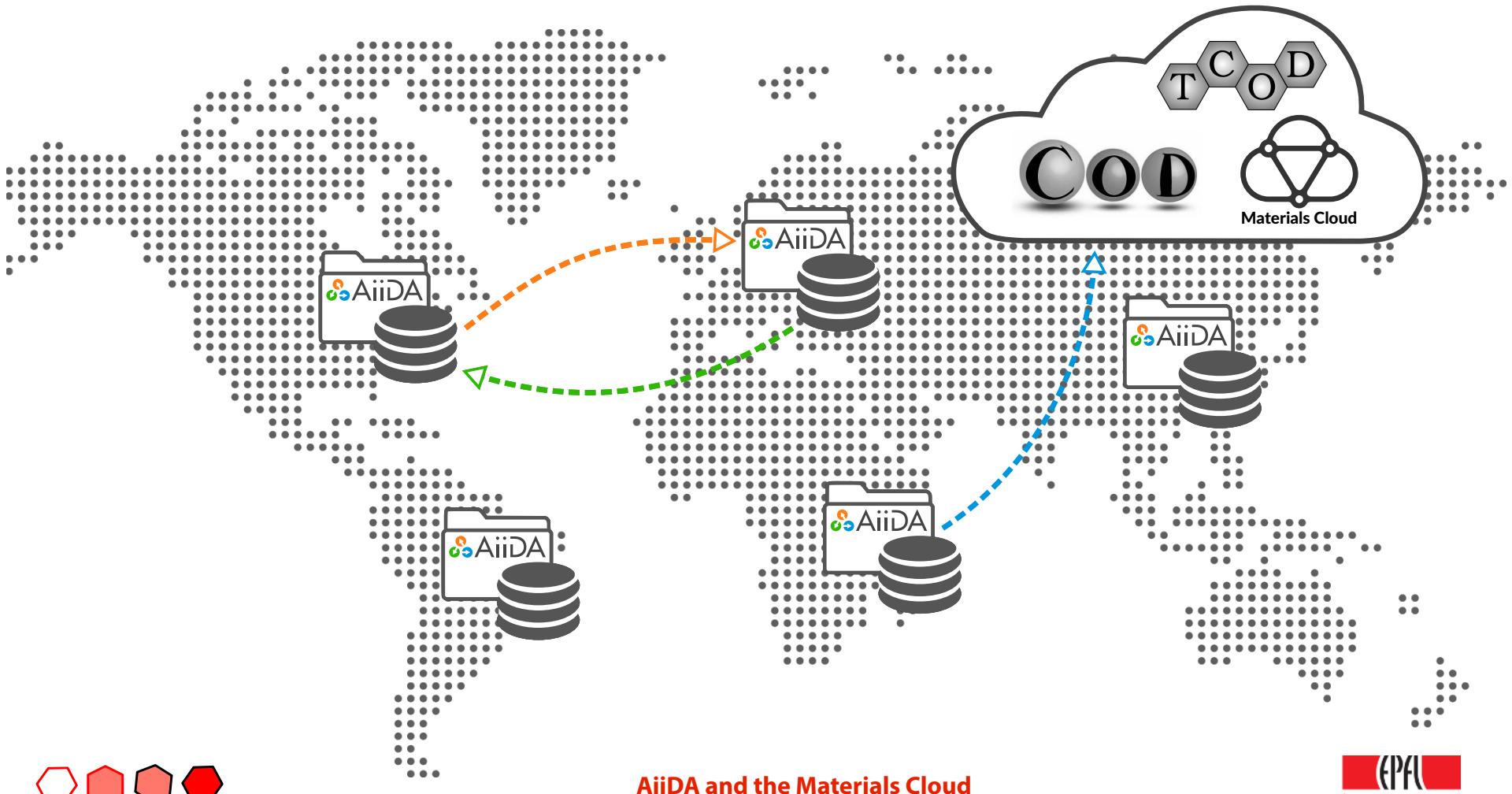
- Turn-key solution:

```
PwBandStructureWorkChain.run(
    code=Code.get_from_string(
        'qe-pw-6.2.1@localhost'),
    structure=StructureData(
        ase=ase.build.bulk('Al')),
    pseudo_family=Str('sssp-pbe-efficiency'))
```



Sharing in AiiDA: data and graphs

- Private AiiDA instances
- UUIDs to uniquely identify nodes
- Data can be shared to other AiiDA repositories or to online repositories



Sharing in AiiDA: codes, plugins and workflows



Calculation



Data



Parsers



Transport and
scheduler



Workflows



Importers &
exporters

AiiDA plugin registry

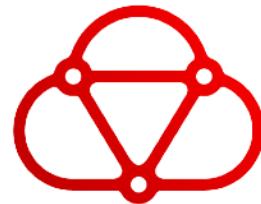
Calculations	71 plugins in 30 entries
Parsers	64 plugins in 30 entries
Data	35 plugins in 18 entries
Workflows	72 plugins in 13 entries
Other	61 plugins in 17 entries

- Plugins are collected in the AiiDA plugin registry
- Over **70 different code executables**
- **currently supported from over 30 different codes,** with almost 50 workflows
- Many are **community-contributed**

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



OPEN SCIENCE PLATFORM:



MATERIALS CLOUD

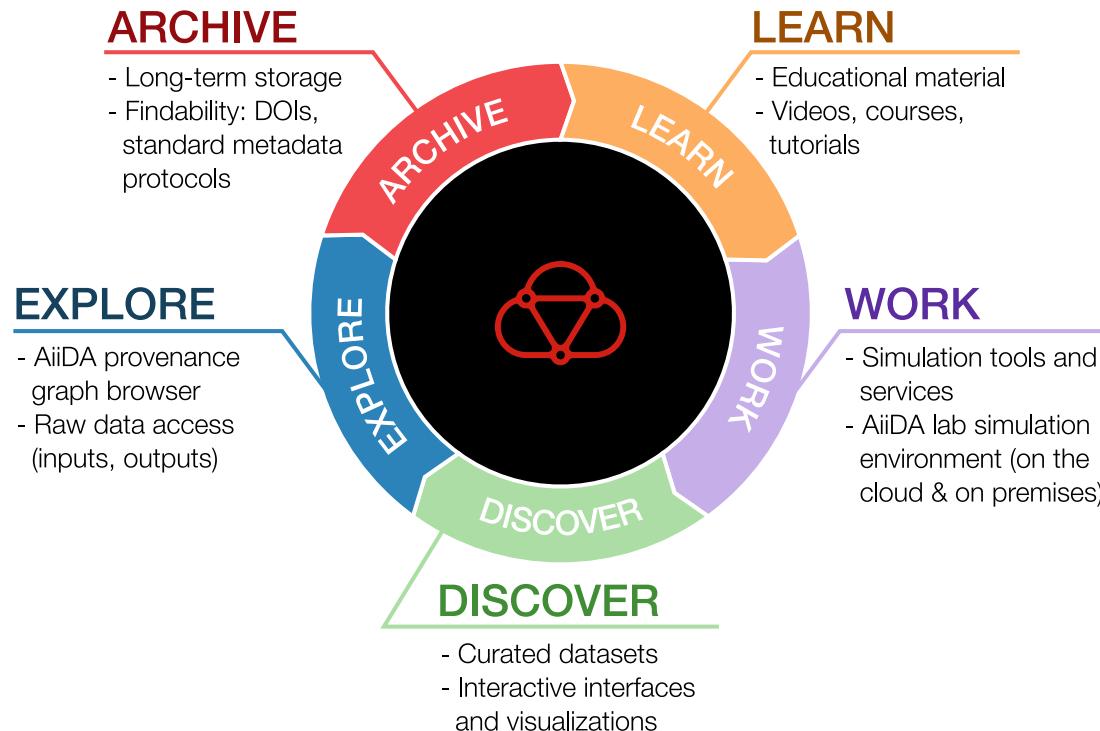
<https://www.materialscloud.org>



AiiDA and the Materials Cloud

Materials Cloud

- **AiiDA** is the 'engine', like **Git** - used in production *since 2015*
- **Materials Cloud** is the dissemination platform (like **GitHub**) **and more** (cloud computing and data generation platform) - online since *Dec 2017*



Data generation: Materials Cloud Work

The screenshot shows the Materials Cloud interface with several service cards:

- Tools**: Online web tools to work with your data.
- AiiDA Lab**: Run your own simulations using AiiDA on the cloud via Jupyter apps.
- Quantum Mobile**: A virtual machine with quantum codes ready to be used via AiiDA. This card is highlighted with a red border.
- AiiDA registry**: The official registry of AiiDA plugins.

AiiDA Lab

- Comes with a preconfigured AiiDA setup, **ideal interface for turn-key workflows**
- Custom **AppMode** extension to make notebooks look&feel like real web apps **knowing only python**
- Using JupyterHub + DockerSpawner



Quantum Mobile

- **Downloadable VM** with preinstalled **AiiDA and codes** like QE, Yambo, Fleur, Siesta, CP2K, ...
- Includes **same AiiDA Lab apps environment** as on Materials Cloud
- Ideal for **education**

AiiDA Lab: submitting a turn-key solution (phonons)

The screenshot shows the AiiDA Lab interface running in a web browser. The top navigation bar includes standard OS X-style buttons (red, yellow, green), a URL field (aiidalab.materialscloud.org), and links for 'Edit App', 'Logout', 'Control Panel', and 'Materials Cloud'. The main content area is organized into sections:

- Home**: Contains icons for 'File Browser' (document icon), 'Terminal' (terminal icon), 'Tasks' (server icon), and 'Manage Apps' (gear icon). There are also up/down arrows for reordering.
- AiiDA**: Lists 'Daemon Status', 'Workflows', 'REST API' on the left, and 'Graph Browser', 'Delete nodes' on the right. Reordering arrows are present.
- Quantum Mobile**: Lists 'Setup Connection to Quantum-Mobile' and 'Setup Codes'. Reordering arrows are present.
- LSMO apps**: Contains four categories with sub-links:
 - Isotherm**: 'Compute one', 'Compute multiple'
 - Charges**: 'Compute Charges'
 - Pore analysis**: 'Compute Pores'
 - Import your data to AiiDA**: 'Import database', 'Plot imported data'Reordering arrows are present.



AiiDA and the Materials Cloud

AiiDA Lab: checking the workflow results (phonons)

The screenshot shows a web browser window for aiidalab.materialscloud.org. The title bar indicates the current page is ".start-15" and the tab is ".phonon turnkey solution-14". The main content area displays a Python code snippet for a workflow configuration:

```
{'automatic_parallelization': {'max_num_machines': 2,
    'max_wall_time_seconds': 86400,
    'target_time_seconds': 3600},
'cutoff_factor': 1.0,
'deg': 0.02,
'do_bands': True,
'do_mag': False,
'do_phonons': True,
'kpts_dist': 0.2,
'matdyn_codename': 'matdyn-6.1@daint-s761',
'mixing_beta': 0.7,
'ph_codename': 'ph-6.1@daint-s746',
'psfam': 'SSSP_eff_PBE',
'pw_codename': 'pw-6.1@daint-s746',
'q2r_codename': 'q2r-6.1@daint-s761',
'relaxation_scheme': 'vc-relax',
'smear': 'cold'}
```

Below the code, there are sections for "structure", "Cell", and "Sites", each containing JSON-like data:

```
structure
<StructureData: uuid: abcab877-1aee-4c8a-85a2-6500a507b578 (unstored)>
Cell
[[0.0, 1.78, 1.78], [1.78, 0.0, 1.78], [1.78, 1.78, 0.0]]
Sites
[<Site: kind name 'C' @ 0.0,0.0,0.0>, <Site: kind name 'C' @ 0.89,0.89,0.89>]
```

A green button labeled "► Run workflow with validated input" is visible. Below it, a message states: "The workflow was run with **workflow ID = 114**. Please note down this number for future reference." A link "Monitor the status and check results" is provided, followed by "Workflow 114".



AiiDA Lab: checking the workflow results (STM)

Jupyter Edit App Logout Control Panel Materials Cloud

Search AiiDA Database for Slab Models

PKs: e.g. 4062 4753 (space separated)

Formulas: e.g. C44H16 C36H4

Calculation Na...: e.g. a great name.



github.com/cpignedoli/mc-empa-surfaces

A solution tailored for every user

User	Skills	Goals	Solution
Computational Scientist	Knows Unix, bash, python	<ul style="list-style-type: none">run high-throughput calculationswrite complex workflowsdevelop AiiDA plugins	AiiDA on the laptop
Experimental Scientist	<i>Doesn't know</i> Unix, bash, python	<ul style="list-style-type: none">run pre-defined workflowsanalyze results	AiiDA Lab in the cloud
Student (tutorial/lecture)	<i>some familiarity with</i> Unix, bash, python	<ul style="list-style-type: none">learn how to use AiiDAlearn how to use ab-initio codestake materials home	Quantum Mobile on the laptop

Open data sharing: Archive, Discover, Explore

DOIs
assigned

materialscloud:2017.0008

SCIENTIFIC DATA



re3data.org
FAIR data
standards, databases, policies

FAIRsharing.org

FAIRsharing.org
re3data.org

+

Recommended
data repository
by Nature's
journal
Scientific Data

DOI [10.24435/materialscloud:2017.0008/v2](https://doi.org/10.24435/materialscloud:2017.0008/v2) (version v2, submitted on 21 March 2018)

How to cite this entry

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 (2018), doi: [10.24435/materialscloud:2017.0008/v2](https://doi.org/10.24435/materialscloud:2017.0008/v2).

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, etc.) 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



AiiDA and the Materials Cloud

Open data sharing: Archive, Discover, Explore

[2017.0008/v1] Two-dimens... x + https://archive.materialscloud.org/2017.0008/v1

LEARN WORK DISCOVER EXPLORE ARCHIVE More

materialscloud:2017.0008

Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds (Data download)

Authors: Nicolas Mounet^{1*}, Marco Gibertini¹, Philippe Schwaller¹, Davide Campi¹, Andrius Merkys^{1,2}, Antimo Marrazzo¹, Thibault Sohier¹, Ivano E. Castelli¹, Andrea Cepellotti¹, Giovanni Pizzi¹, 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, Saulėtekio al. 7, LT-10257 Vilnius, Lithuania
* Corresponding authors emails: nicolas.mounet@epfl.ch, nicola.marzari@epfl.ch

(version: v1, submitted on: 21 December 2017)
[There are newer versions. Click here to view the latest version v2]

How to cite this entry DOI 10.24435/materialscloud:2017.0008/v1

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 (Data download)*, Materials Cloud Archive (2017), doi: 10.24435/materialscloud:2017.0008/v1.

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, etc.) 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	111.1 MiB	We provide 258 two-dimensional crystal structures (lattice vectors, atomic



AiiDA and the Materials Cloud

DATA MANAGEMENT PLANS 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, reproduce in the AiiDA Lab thanks to *full provenance*
- We provide **DMP templates** for researchers using Materials Cloud (Swiss SNSF, EU H2020, and more in the future)

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

Funding Body

DMP template (using  AiiDA)

DMP template (no AiiDA)

SNF

.docx

.odt

.pdf

.docx

.odt

.pdf



Acknowledgements and funding



swissuniversities



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

EPFL Open Science Fund “OSSCAR”

Creating a hub of computational resources geared
also towards education and teaching

Moreover: *H2020 Marketplace* (providing data and simulation services in a EU Marketplace platform also for industry); *H2020 Intersect* (develop AiiDA workflows to compute transport properties of materials)



AiiDA and the Materials Cloud



The Materials Cloud And AiiDA teams

Acknowledgements



Marco
Borelli
(EPFL)



Sebastiaan
P. Huber
(EPFL)



Boris
Kozinsky
(BOSCH)



Snehal P.
Kumbhar
(EPFL)



Leonid
Kahle
(EPFL)



Nicola
Marzari
(EPFL)



Elsa
Passaro
(EPFL)



Giovanni
Pizzi
(EPFL)



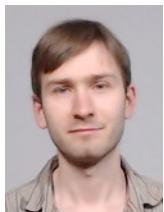
Thomas
Schulthess
(ETHZ,CSCS)



Berend
Smit
(EPFL)



Leopold
Talirz
(EPFL)



Martin
Uhrin
(EPFL)



Joost
VandeVondele
(ETHZ,CSCS)



Aliaksandr
Yakutovich
(EPFL)



Spyros
Zoupanos
(EPFL)

Contributors for the 20+ plugins: **Quantum ESPRESSO, Wannier90, CP2K, FLEUR, YAMBO, SIESTA, VASP, ...**

Contributors to `aiida_core` and former AiiDA team members — Valentin Bersier, Jocelyn Boullier, Jens Broeder, Andrea Cepellotti, Fernando Gargiulo, Christoph Koch, Dominik Gresch, Rico Häselmann, Eric Hontz, Andrius Merkys, Nicolas Mounet, Tiziano Müller, Ole Schütt, Riccardo Sabatini, Phillippe Schwaller

The CSCS support teams

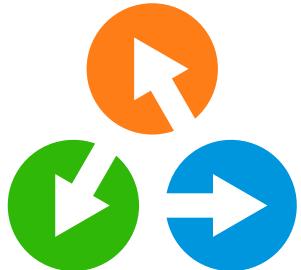
Summary

- **AiiDA**: the reproducibility and automation engine
 - Define turn-key workflows, automate them on supercomputers
 - Keep track of the **provenance** in the form of a graph
 - Share data, plugins and workflows
- **Materials Cloud**: the dissemination portal
 - **AiiDA Lab**: data generation platform, both on the cloud or locally (Quantum Mobile VM)
 - *Archive*: Findable long-term storage + *Discover* (curated data) + *Explore* (raw AiiDA data): FAIR sharing
 - Supports researchers in data management and DMPs

Open Science: not only *open data, codes and workflows*,
but also ***straightforward access to them***



Contacts



Website: <http://www.aiida.net>

Docs: <http://aiida-core.readthedocs.io>

Git repo: https://github.com/aiidateam/aiida_core/

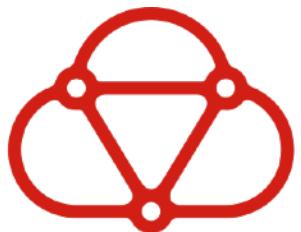
Plugin registry: <http://aiidateam.github.io/aiida-registry>



<https://www.facebook.com/aiidateam>



@aiidateam



Materials Cloud: <http://www.materialscloud.org>

- **AiiDA Lab:** <http://aiidalab.materialscloud.org>
- **Archive:** <http://archive.materialscloud.org>

Quantum Mobile: <http://www.materialscloud.org/work/quantum-mobile>