



# Generating and publishing datasets using AiiDA and Materials Cloud

**Kristjan Eimre**  
**EPFL, Switzerland**

CECAM workshop: Machine Learning Interatomic Potentials and Accessible Databases  
Grenoble  
11.09.2024

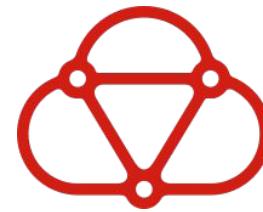
# Outline

---



Automated Interactive Infrastructure and  
Database for Computational Science

- Recent usability improvements



Web platform for seamlessly sharing data  
and resources

- Materials Cloud 3D crystals database updates
- Archive-OPTIMADE integration



---

<http://www.aiida.net>

Computational science infrastructure that provides

- scalable workflow engine
- built-in support for HPC
- flexible plugin system
- automatic full data provenance

Language: implemented and API in python

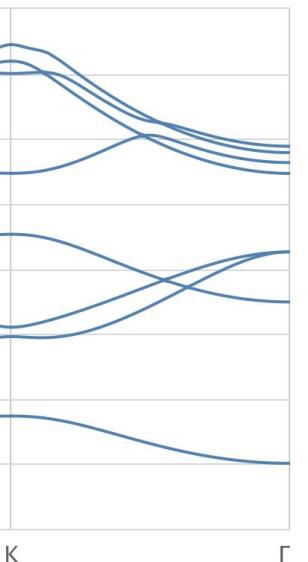
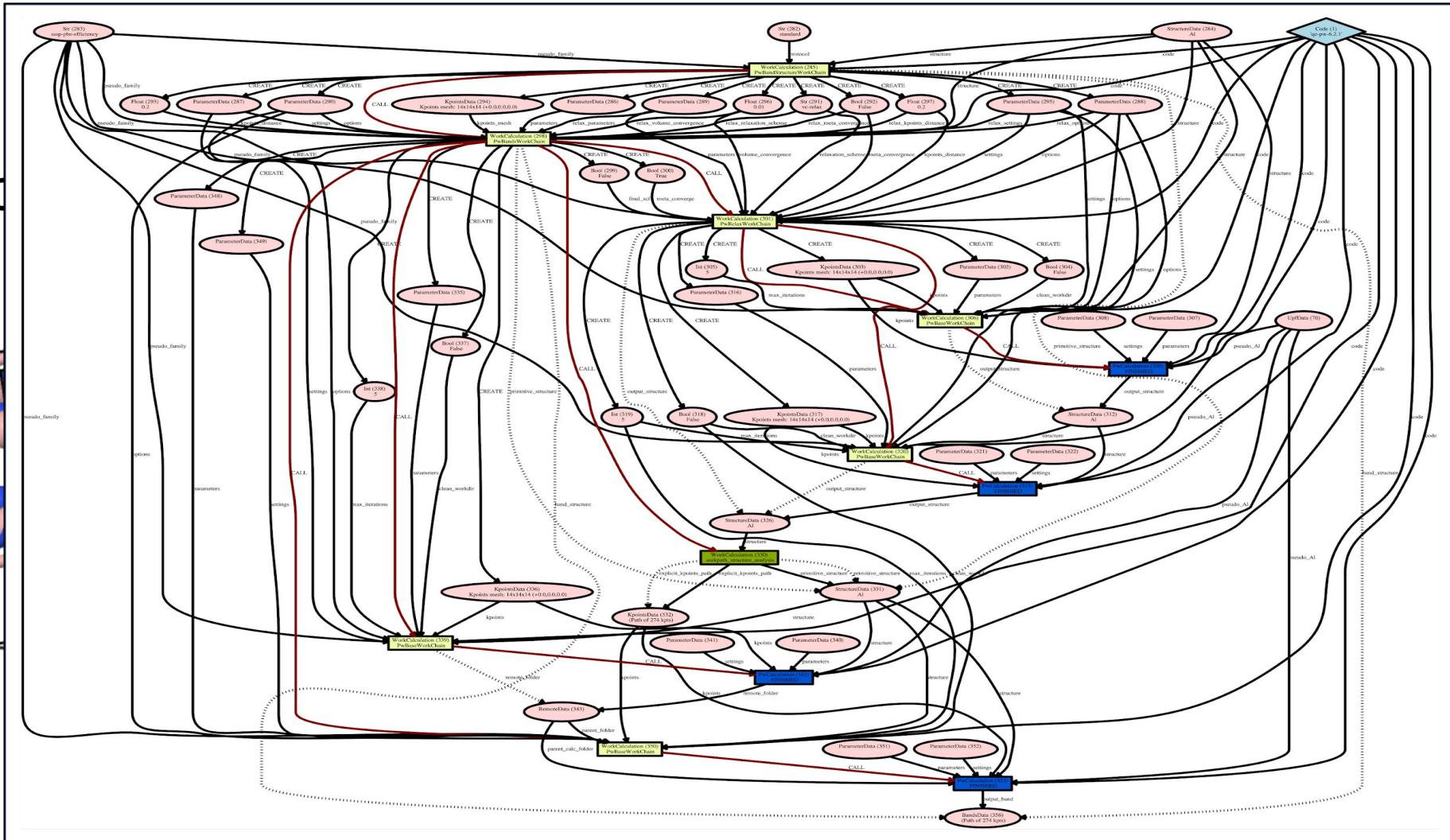
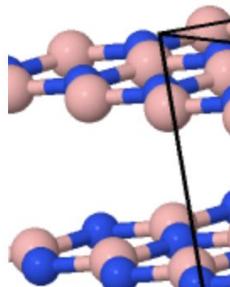
Open source: <https://github.com/aiidateam/aiida-core>

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

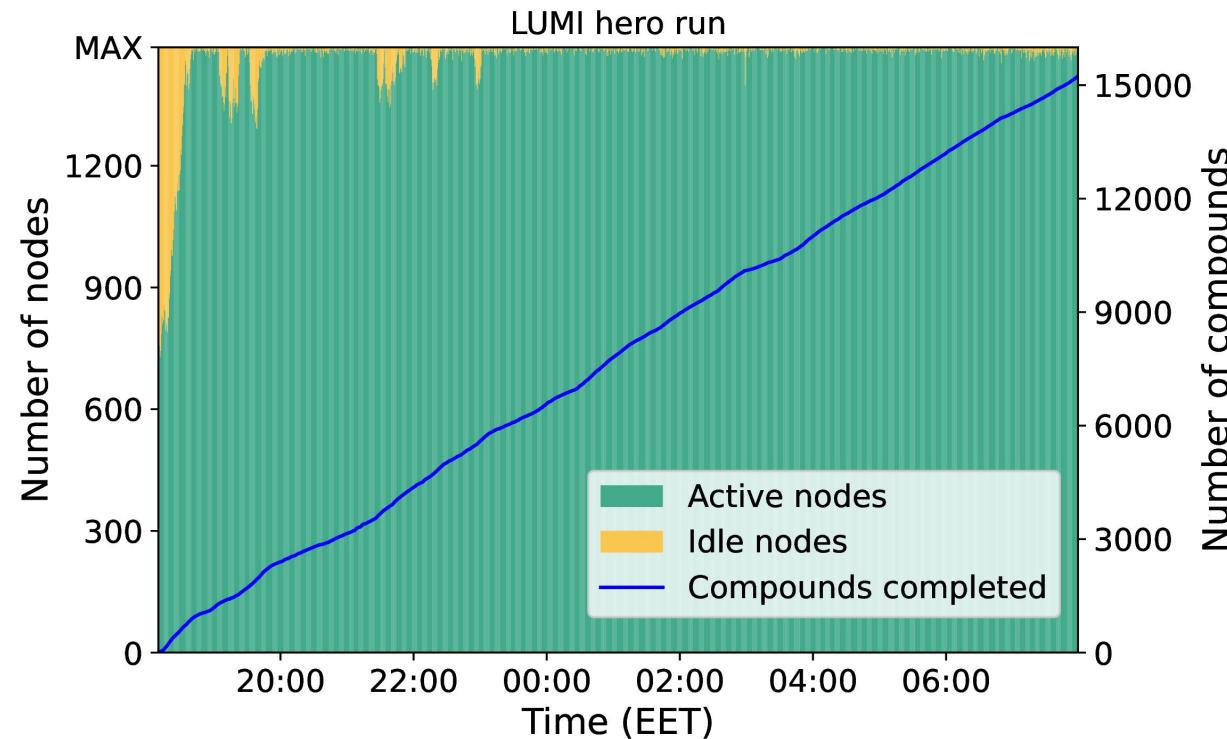
S.P. Huber et al., Scientific Data 7, 300 (2020)

# Workflow engine

# Example



## Scalable workflow engine

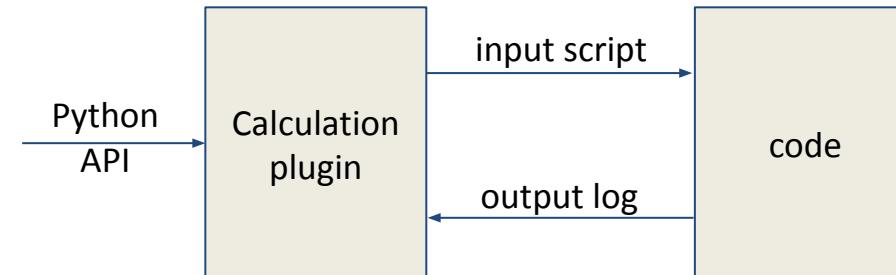


- Geometry relaxations of 15,324 inorganic compounds in 13 hours
- Full partition of LUMI-C: 1,500 nodes with 128 cores each
- 7887 issues dealt with on the fly

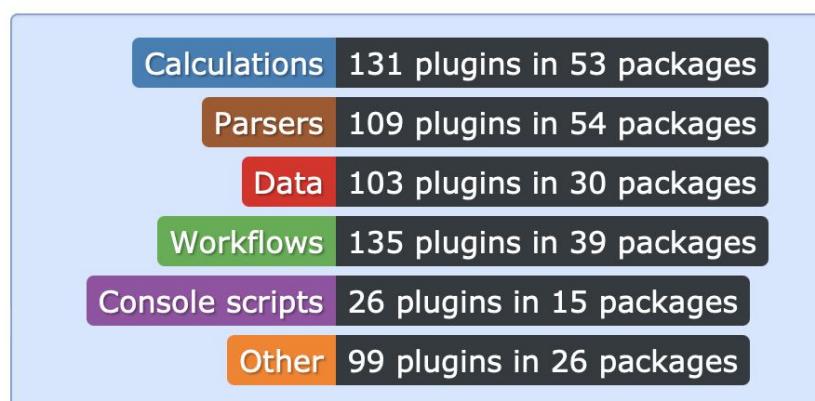
# Plugins

---

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



**Registered plugin packages: 91**



- Plugins collected in the AiiDA plugin registry
- 130+ codes currently supported
- Many are community-contributed

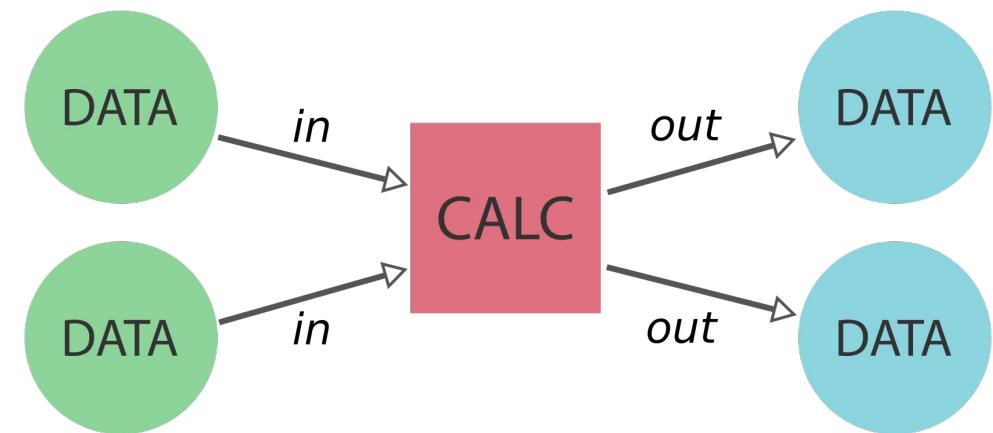
# Data provenance

---

What you get as a bonus: full data provenance!

AiiDA stores:

- data transformations/calculations
- inputs
- outputs
- inter-connections



# Data provenance

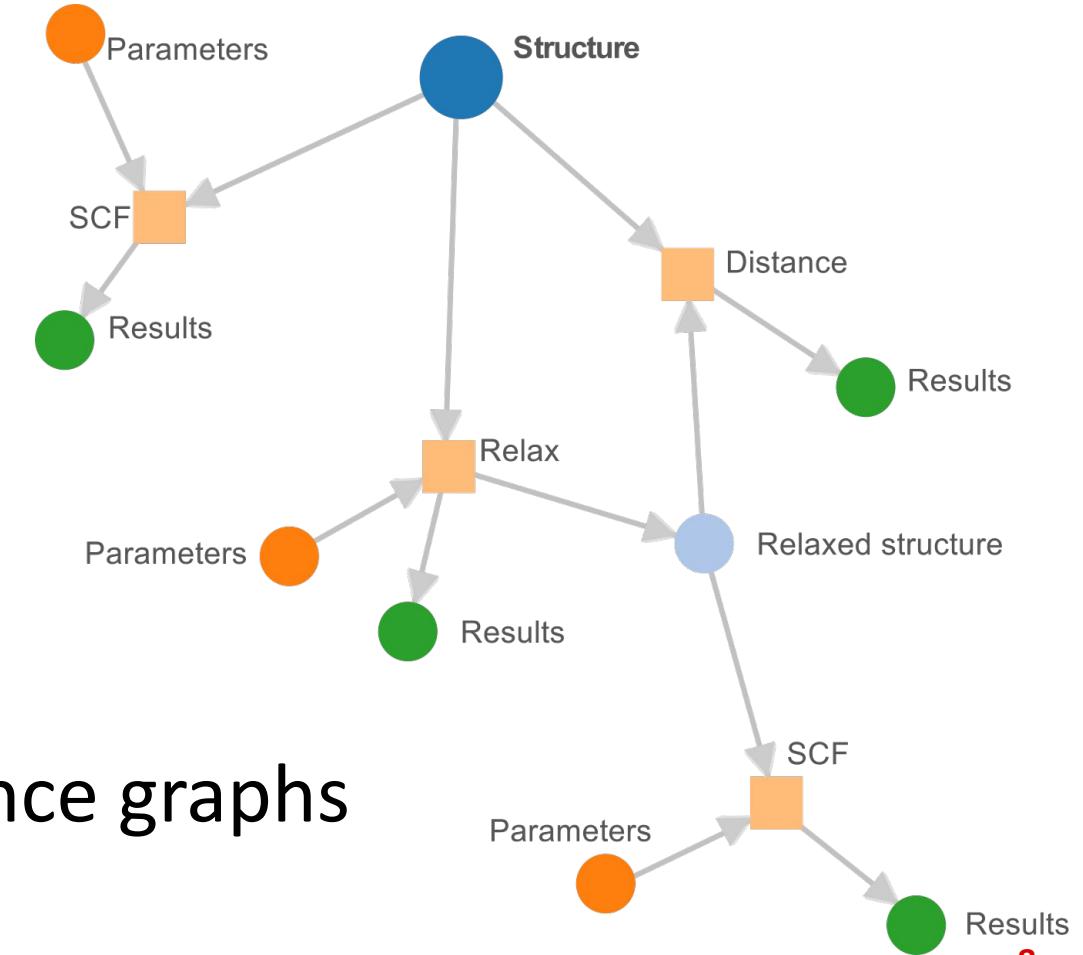
---

What you get as a bonus: full data provenance!

AiiDA stores:

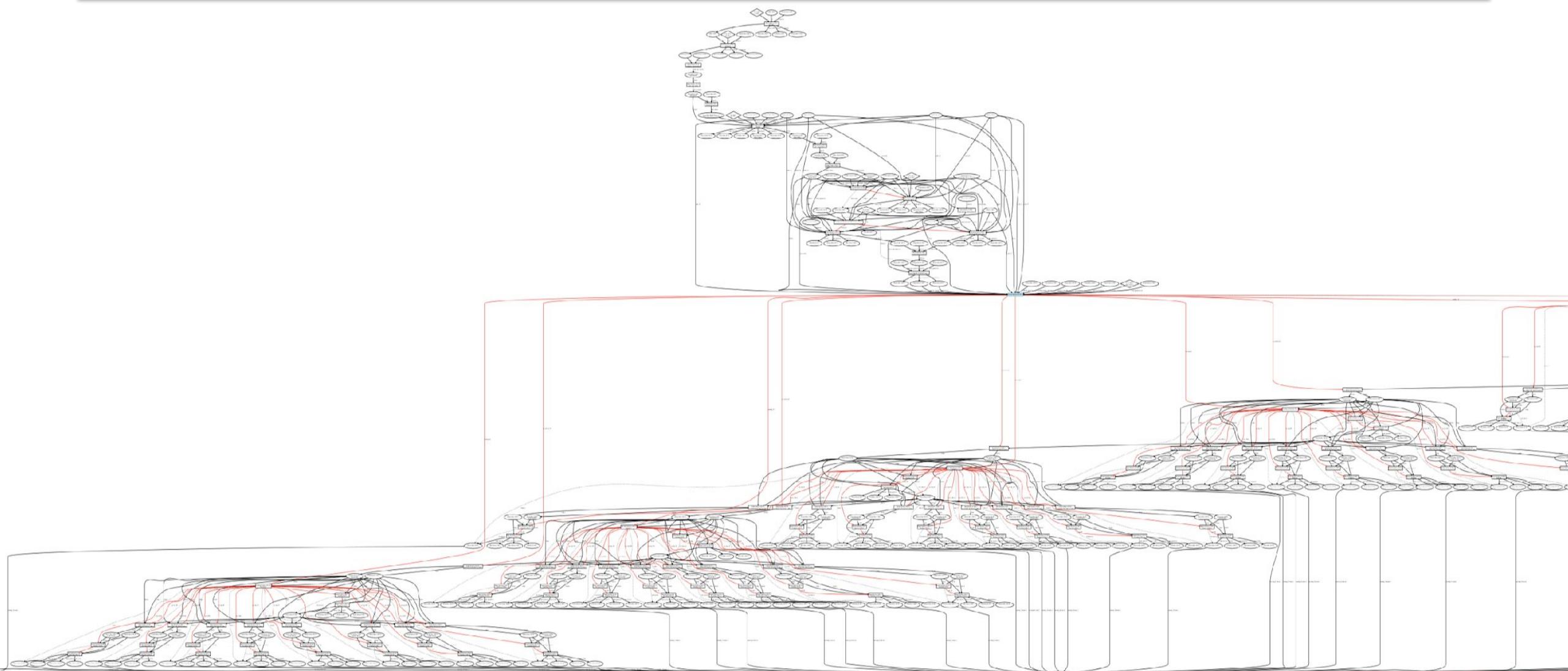
- data transformations/calculations
- inputs
- outputs
- inter-connections

Provenance graphs



# Data provenance

---



Molecular dynamics study of Lithium in a solid electrolyte

# AiiDA improvements

---

Recent effort to improve the usability and reduce  
the learning curve

# AiiDA improvements: installation

## Traditional (Complete) installation

- Requires
  - PostgreSQL
  - RabbitMQ

```
$ pip install -q aiida-core
$ verdi profile setup core.psql_dos
Report: enter ? for help.
Report: enter ! to ignore the default and set no
Profile name: new_profile
Set as default? [Y/n]: n
Email Address (for sharing data) [kristjan.eimre@epfl.ch]:
First name [Kristjan]:
Last name [Eimre]:
Institution [EPFL]:
Use RabbitMQ? [Y/n]:
PostgreSQL engine [postgresql_psycopg2]:
PostgreSQL hostname [localhost]:
PostgreSQL port [5432]:
PostgreSQL username:
```

## verdi presto (v2.6.1, July 1st)

```
$ pip install -q aiida-core
$ verdi presto
Report: Option `--use-postgres` not enabled: config
Report: RabbitMQ server not found (Failed to connect)
Report: See https://aiida-core.readthedocs.io/en/stable/installation/storage.html#storage
Report: Initialising the storage backend.
Report: Storage initialisation completed.
Success: Created new profile `presto-3`.
Success: Configured the localhost as a computer.
$ verdi status
✓ version:      AiiDA v2.6.2
✓ config:        /home/kristjan/.aiida
✓ profile:       presto-3
✓ storage:       SqliteDosStorage[/home/kristjan/.aiida]
● broker:        No broker defined for this profile:
● daemon:        No broker defined for this profile:
$
```

Not suitable for high-throughput

# AiiDA improvements: aiida-shell

**aiida-shell:** run any executable through AiiDA without a plugin

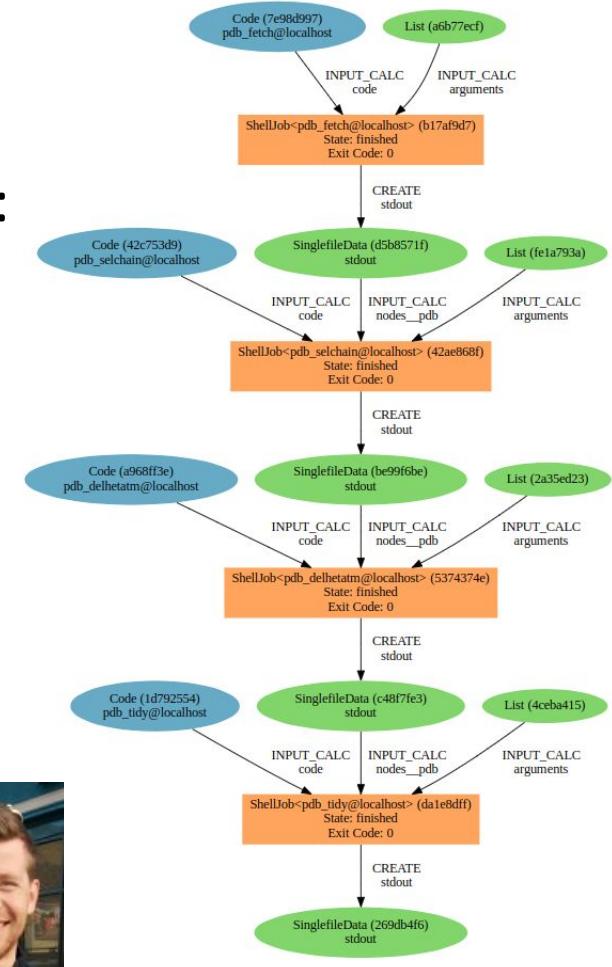
Example shell commands (manipulating Protein Data Bank files):

```
$ pdb_fetch 1brs | pdb_selchain -A,D | pdb_delhetatm | pdb_tidy > 1brs_AD_noHET.pdb
```

Run via aiida-shell:

```
from aiida_shell import launch_shell_job

results, node = launch_shell_job('pdb_fetch', '1brs')
results, node = launch_shell_job('pdb_selchain', '-A,D {pdb}', {'pdb': results['stdout']})
results, node = launch_shell_job('pdb_delhetatm', '{pdb}', {'pdb': results['stdout']})
results, node = launch_shell_job('pdb_tidy', '{pdb}', {'pdb': results['stdout']})
```



<https://github.com/sphuber/aiida-shell>

Dr. Sebastiaan  
P. Huber

# AiiDA improvements: WorkGraph

Workflow:  $(x + y) * z$  extending to  $(x + y) * z + a$

## Calculations

```
@calcfunction
def add(x, y):
    return Int(x + y)

@calcfunction
def multiply(x, y):
    return Int(x * y)
```

## WorkChain

```
class AddMultiplyWorkChain(WorkChain):

    @classmethod
    def define(cls, spec):
        super().define(spec)
        spec.input('x')
        spec.input('y')
        spec.input('z')
        spec.outline(
            cls.add,
            cls.multiply,
            cls.results,
        )
        spec.output('result')

    def add(self):
        self.ctx.sum = add(self.inputs.x,
self.inputs.y)

    def multiply(self):
        self.ctx.product =
multiply(self.ctx.sum, self.inputs.z)

    def results(self):
        self.out('result', self.ctx.product)
```

## WorkGraph

```
wg = WorkGraph("add_multiply")
wg.tasks.new(add, name="add")
wg.tasks.new(multiply, name="multiply",
x=wg.tasks["add"].outputs["result"])
```

<https://github.com/aiidateam/aiida-workgraph>



Dr. Xing  
Wang

# AiiDA improvements: WorkGraph

Workflow:  $(x + y) * z$  extending to  $(x + y) * z + a$

```
class AddMultiplyAddWorkChain(WorkChain):  
  
    @classmethod  
    def define(cls, spec):  
        super().define(spec)  
        spec.expose_inputs(AddMultiplyWorkChain,  
                           namespace='add_multiply')  
        spec.input('a')  
        spec.outline(  
            cls.add_multiply,  
            cls.add2,  
            cls.results,  
        )  
        spec.output('result')  
  
    def add_multiply(self):  
        future =  
self.submit(AddMultiplyWorkChain,  
**self.exposed_inputs(AddMultiplyWorkChain,  
'add_multiply'))  
        return ToContext(add_multiply=future)  
  
    def add2(self):  
        self.ctx.sum =  
add(self.ctx.add_multiply.outputs.result,  
self.inputs.a)  
  
    def results(self):  
        self.out('result', self.ctx.sum)
```

## WorkChain

```
class AddMultiplyWorkChain(WorkChain):  
  
    @classmethod  
    def define(cls, spec):  
        super().define(spec)  
        spec.input('x')  
        spec.input('y')  
        spec.input('z')  
        spec.outline(  
            cls.add,  
            cls.multiply,  
            cls.results,  
        )  
        spec.output('result')  
  
    def add(self):  
        self.ctx.sum = add(self.inputs.x,  
self.inputs.y)  
  
    def multiply(self):  
        self.ctx.product =  
multiply(self.ctx.sum, self.inputs.z)  
  
    def results(self):  
        self.out('result', self.ctx.product)
```

## WorkGraph

```
wg = WorkGraph("add_multiply")  
wg.tasks.new(add, name="add")  
wg.tasks.new(multiply, name="multiply",  
x=wg.tasks["add"].outputs["result"])  
  
wg.tasks.new(add2, name="add2",  
x=wg.tasks["multiply"].outputs["result"])
```

<https://github.com/aiidateam/aiida-workgraph>



Dr. Xing  
Wang

# AiiDA improvements

---

- verdi storage backup <destination>
  - incremental backup of the AiiDA profile
  - supports remote paths directly
- verdi process dump <uuid/pk>
  - dumps all input and output scripts of the selected aiida process into a human-readable folder structure

```
> verdi process dump 7727b19f-5c77-4d2a-bbb0-71a8cf87b4b7
Success: Raw files for WorkChainNode <485293> dumped into
> cd dump-PwBaseWorkChain-485293/
> tree
:
├── 01-create_kpoints_from_distance
│   └── inputs
│       └── source_file
├── 02-iteration_01-PwCalculation
│   ├── inputs
│   │   └── _aiidasubmit.sh
│   └── aiida.in
└── node_inputs
    └── pseudos
        ├── Ge
        │   └── ge_pbesol_v1.4.uspp.F.UPF
        └── P
            └── P.pbesol-n-rrkjus_ps1.1.0.0.UPF
    └── outputs
        ├── _scheduler-stderr.txt
        ├── _scheduler-stdout.txt
        └── aiida.out
    └── data-file-schema.xml
└── README.md

9 directories, 10 files
```

# AiiDA discourse forum

<https://aiida.discourse.group/>

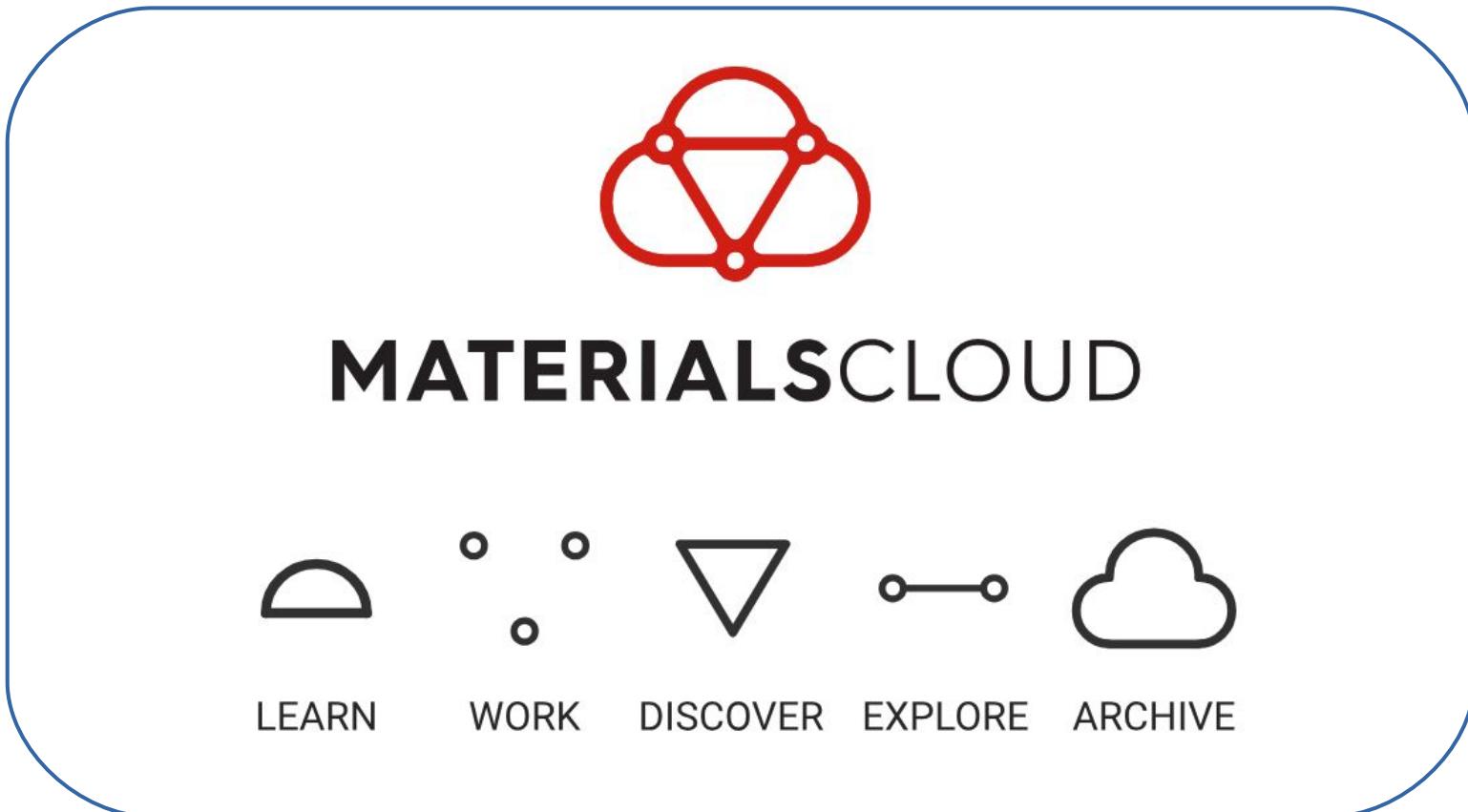
The screenshot shows the homepage of the AiiDA discourse forum. The left sidebar contains navigation links for Topics, My Posts, More, Categories, Tags, question, aiida, plugin, discussion, aiidalab, All tags, Messages, Inbox, and Channels. The main content area features a navigation bar with categories, tags, Latest (selected), Unread (7), Hot, and Categories, followed by a '+ New Topic' button. Below this is a table listing five topics:

Topic	Replies	Views	Activity
Run only one job on local machine Developer question, aiida, plugin, schedulers	1	11	1d
Pass curly braces literally in aiida-shell arguments General Usage plugin	3	9	2d
Aiida-workgraph: How to handle dynamic inputs and outputs of tasks General Usage	10	52	8d
Automatically storing info about Python environment in nodes for calcfunctions General Usage	9	43	9d
EuroSciPy 2024 Support Community	0	19	11d

# Materials Cloud

---

<https://www.materialscloud.org/>



# Discover and explore

## Discover: curated web interface

The screenshot shows the 'Discover' section of the Materials Cloud web interface. At the top, there are navigation tabs: LEARN, WORK, **DISCOVER**, EXPLORE, ARCHIVE, and a 'More' dropdown. Below the tabs, a search bar contains the word 'Discover'. The main content area is titled 'Discover curated data sets' and describes curated research data sets with tailored visualizations maintained by the Materials Cloud team. It features two data entries:

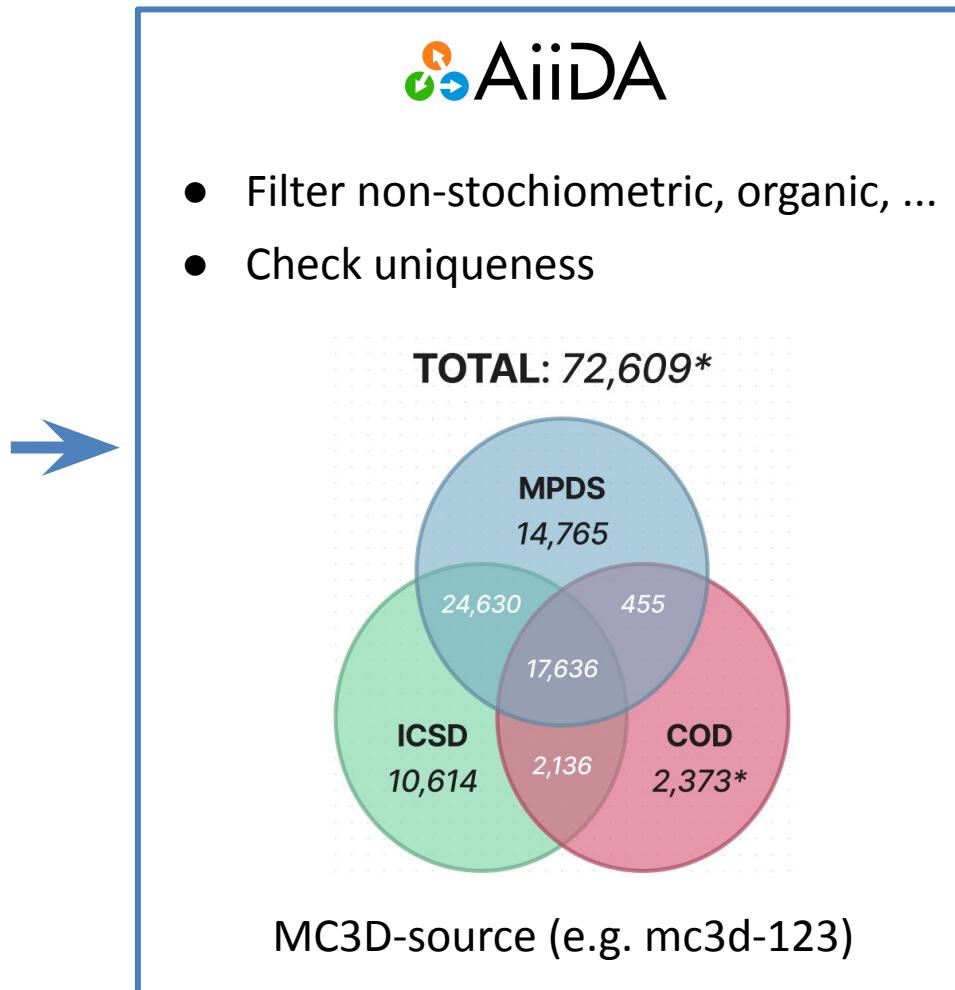
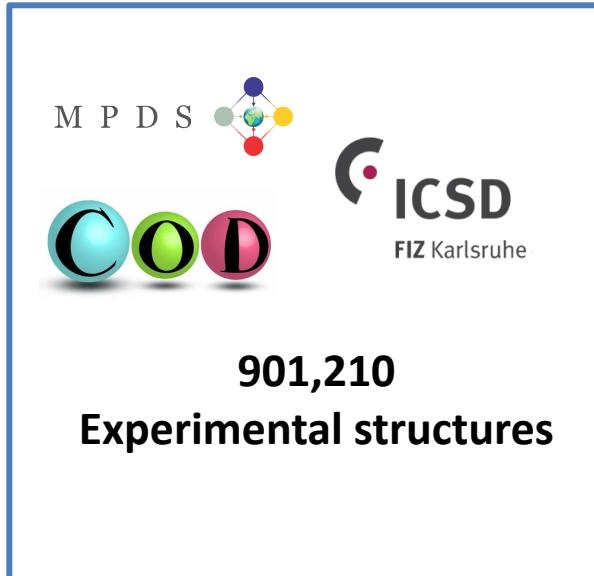
- Materials Cloud three-dimensional crystals database (MC3D)** [DOI 10.24435/materialscloud:rw-t0](#)  
**Authors:** Sebastiaan Huber, Marnik Bercx, Kristjan Eimre, Nicolas Hörmann, Martin Uhrin, Nicola Marzari, Giovanni Pizzi  
**Description:** Curated set of relaxed three-dimensional crystal structures based on raw CIF data from the experimental databases MPDS, COD, and ICSD.
- Materials Cloud two-dimensional crystals database (MC2D)** [DOI 10.24435/materialscloud:az-b2](#)  
[DOI 10.24435/materialscloud:36-nd](#)  
**Authors:** Davide Campi, Davide Grassano, Nicolas Mounet, Marco Gibertini, Philippe Schwaller, Andrius Merkys, Antimo Marrazzo, Thibault Sohier, Ivano E. Castelli, Andrea Cepellotti, Snehal Kumbhar, Elsa Passaro, Kristjan Eimre, Giovanni Pizzi, Nicola Marzari  
**Description:** Results from screening known 3D crystal structures finding those that can be computationally exfoliated, producing 2D materials candidates.

## Explore: visualize AiiDA provenance

The screenshot shows the 'Explore' section of the Materials Cloud web interface. At the top, there are navigation tabs: LEARN, WORK, DISCOVER, **EXPLORE**, ARCHIVE, and a 'More' dropdown. Below the tabs, a search bar contains the word 'Explore'. The main content area is titled 'Explore the full provenance' and describes an interactive browser for exploring AiiDA provenance graphs uploaded to the Materials Cloud Archive. It features two data entries:

- Materials Cloud three-dimensional crystals database (MC3D) (PBE-v1)** [DOI 10.24435/materialscloud:rw-t0](#)  
**Authors:** Sebastiaan Huber, Marnik Bercx, Nicolas Hörmann, Martin Uhrin, Nicola Marzari, Giovanni Pizzi  
**Description:** Curated set of relaxed three-dimensional crystal structures based on raw CIF data from the experimental databases MPDS, COD, and ICSD. Calculations are performed with the PBE-v1 methodology.
- Materials Cloud two-dimensional crystals database (MC2D) (PBEsol-v1)** [DOI 10.24435/materialscloud:az-b2](#)  
[DOI 10.24435/materialscloud:36-nd](#)  
**Authors:** Davide Campi, Davide Grassano, Nicolas Mounet, Marco Gibertini, Philippe Schwaller, Andrius Merkys, Antimo Marrazzo, Thibault Sohier, Ivano E. Castelli, Andrea Cepellotti, Snehal Kumbhar, Elsa Passaro, Kristjan Eimre, Giovanni Pizzi, Nicola Marzari  
**Description:** Curated set of relaxed three-dimensional crystal structures based on raw CIF data from the experimental databases MPDS, COD, and ICSD. Calculations are performed with the PBEsol-v1 methodology.

# Materials Cloud 3D crystals database



MC3D/PBE-v1

Relaxation with 'PBE-v1' methodology.  
34,487 relaxed structures

MC3D/PBEsol-v1

Relaxation with 'PBEsol-v1' methodology.  
33,674 relaxed structures

Properties/screening:

- thermodyn stability
- spectroscopies
- electrides
- superconductors

# Materials Cloud 3D crystals database

<https://mc3d.materialscloud.org/>

Use    About    REST API

Select a methodology: PBEsol-v1    ?  
PBEsol-v1    PBE-v1

Elements filtering mode:  
 Include/exclude     Only selected    ?

ID	Formula	Num. of atoms/cell	Space group number	Is source theoretical?	Is source high pressure?	Is source high temperature?	Total magnetization ( $\mu\text{B}/\text{cell}$ )
mc3d-10/pbesol-v1	Ge <sub>26</sub> Ir <sub>8</sub> Y <sub>6</sub>	40	223	no	no	no	-
mc3d-10000/pbesol-v1	FeSbV	3	216	no	no	no	-
mc3d-10004/pbesol-v1	Cs <sub>2</sub> F <sub>6</sub> Fe	10	225	no	no	yes	5.00
mc3d-10023/pbesol-v1	Ga <sub>12</sub> Mn <sub>4</sub> Zr <sub>6</sub>	22	59	no	no	no	7.09
mc3d-10030/pbesol-v1	B <sub>7</sub> Ca <sub>9</sub> F <sub>2</sub> Li <sub>5</sub> O <sub>21</sub>	44	1	no	no	no	-
mc3d-10031/pbesol-v1	C <sub>4</sub> N <sub>6</sub> Se <sub>4</sub>	16	62	no	no	no	-
mc3d-10044/pbesol-v1	Nb <sub>12</sub> Ni <sub>6</sub> Si <sub>4</sub>	24	227	no	no	no	-
mc3d-10048/pbesol-v1	Ca <sub>6</sub> Ga <sub>4</sub> P <sub>8</sub>	18	15	no	no	no	-
mc3d-10049/pbesol-v1	Cs <sub>4</sub> F <sub>20</sub> O <sub>2</sub> Sb <sub>4</sub>	30	13	no	no	no	-
mc3d-1005/pbesol-v1	H <sub>8</sub> O <sub>18</sub> P <sub>4</sub> Rh <sub>2</sub>	32	5	no	no	no	3.96

Showing 33674 entries out of 33674    ?

Reset column filters    Show columns

Page Size: 20    1 to 20 of 33,674    < < Page 1 of 1,684 > >

Download filtered entries    ?

# Materials Cloud 3D crystals database

<https://mc3d.materialscloud.org/>

Use   About   REST API

Select a methodology: PBESol-v1   [?](#)

Elements filtering mode:  
 Include/exclude   [?](#)  
 Only selected

Showing 9 entries out of 33674   [?](#)

ID	Formula	Num. of atoms/cell	Space group number	Is source theoretical?	Is source high pressure?	Is source high temperature?	Total magnetization (μB/cell)
mc3d-10103/pbesol-v1	CoNi	2	221	yes	no	no	
mc3d-22117/pbesol-v1	Co <sub>2</sub> Li <sub>9</sub> Mn <sub>3</sub> Ni <sub>3</sub> O <sub>18</sub>	36	151	no	no	no	
mc3d-31847/pbesol-v1	Co <sub>2</sub> GaNi	4	139	yes	no	no	
mc3d-3377/pbesol-v1	Co <sub>6</sub> Ni <sub>3</sub> V <sub>3</sub>	12	166	no	no	no	
mc3d-40385/pbesol-v1	Co <sub>3</sub> Ni	4	221	yes	no	no	
mc3d-43549/pbesol-v1	Co <sub>6</sub> F <sub>6</sub> Ni	8	148	no	no	no	
mc3d-43590/pbesol-v1	Co <sub>2</sub> GaNi	4	225	yes	no	no	
mc3d-69079/pbesol-v1	CoGeMnNi	4	216	no	no	no	4.63
mc3d-72144/pbesol-v1	Co <sub>2</sub> GaNi	4	123	no	no	no	3.02

Reset column filters   Show columns

Total magnetization of the ferromagnetic solution, if it was found.  
Apply filter:  
Greater than  AND  OR  
Equals   
Reset   Apply

Page Size: 20   1 to 9 of 9   < Page 1 of 1 > >

Download filtered entries   [?](#)

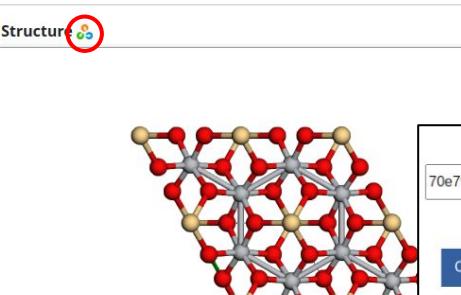
# Materials Cloud 3D crystals database

<https://mc3d.materialscloud.org/>

Materials Cloud three-dimensional crystals database (MC3D) DOI: 10.24435/materialscloud:rw-t0

CdO<sub>6</sub>Ti<sub>2</sub> (mc3d-10114/pbesol-v1)

General overview

Structure 

Info  
Formula: CdO<sub>6</sub>Ti<sub>2</sub>

CalcJobNode  
JOB ID: 308095  
SCHEDULER STATE: done  
REMOTE WORKING DIRECTORY: /scratch/e1000/mbercx/aiida/70e7946c-ae04-42b5-a3c8-87628be2594c

X-ray diffraction pattern  
Select the X-ray source: CuK $\alpha$ , Select peak broadening profile: Gaussian

Provenance links  
Relevant nodes in the provenance browser:

- Final optimization calculation
- Final structure

AIIDA Provenance Browser  
Click on node to browse, drag to animate

Structural details

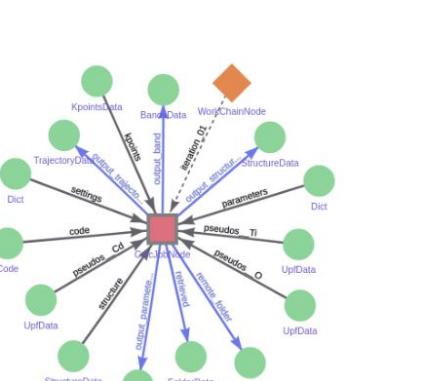
General  
Explore provenance   
Download structure 

Cell

	x [Å]	y [Å]
V <sub>1</sub>	5.1469	0.0000
V <sub>2</sub>	-2.5735	4.4574
V <sub>3</sub>	0.0000	0.0000

Last Job Information

UUID: 70e7946c-ae04-42b5-a3c8-87628be2594c  
Type: aiida.calculations:quantumespresso.pw  
Created on 24 September 2021  
Modified 10 months ago  
Creator: Marnik Berckx (EPFL)



# Materials Cloud 3D crystals database

## Accessing data directly

**Materials Cloud three-dimensional crystals database (MC3D)**

DOI 10.24435/materialscloud:rw-t0



Materials Cloud three-dimensional crystals database is a curated set of computationally relaxed three-dimensional crystal structures and calculated properties. The crystal structures originate from experimental databases.

[Use](#) [About](#) [REST API](#)

This section contains an overview of our REST APIs to access the MC3D data.

### 1. Materials Cloud and AiiDA REST APIs

The MC3D frontend is running on the following APIs:

- Index of materials: <https://dev-aiida.materialscloud.org/mc-rest-api/mc3d/pbe-v1/entries>
- Single entry data: <https://dev-aiida.materialscloud.org/mc-rest-api/mc3d/pbe-v1/entries/mc3d-10>
- AiiDA REST API for properties and provenance: <https://dev-aiida.materialscloud.org/mc3d/api/v4>

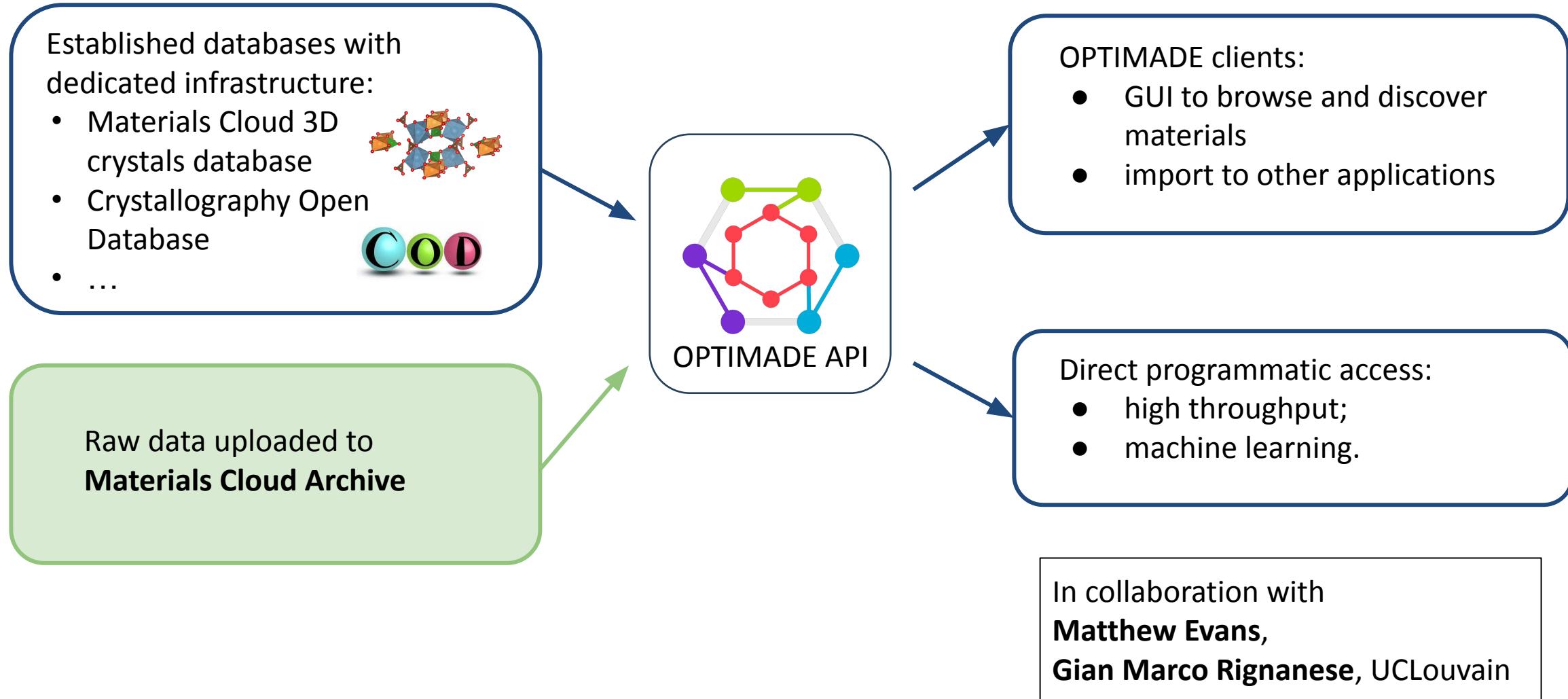
### 2. OPTIMADE REST API

The MC3D database can also be accessed via an API following the [OPTIMADE specification](#). This currently only includes the crystal structures and no properties or provenance information is provided. Relevant endpoints are

- <https://aiida.materialscloud.org/mc3d/optimade/v1/info>
- <https://aiida.materialscloud.org/mc3d/optimade/v1/structures>

23

# Archive-OPTIMADE service



# Archive-OPTIMADE service

<https://archive.materialscloud.org>

## Machine learning-accelerated discovery of A<sub>2</sub>BC<sub>2</sub> ternary electrides with diverse anionic electron densities

Zhiqi Wang<sup>1</sup>, Yutong Gong<sup>1</sup>, Matthew L. Evans<sup>2\*</sup>, Yujing Yan<sup>1</sup>, Shiyao Wang<sup>1</sup>, Nanxi Miao<sup>1</sup>, Ruiheng Zheng<sup>1</sup>, Gian-Marco Rignanese<sup>1,2\*</sup>, Junjie Wang<sup>1\*</sup>

1 State Key Laboratory of Solidification Processing, School of Materials Science and Engineering, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, People's Republic of China.

2 IMCN-MODL, Université catholique de Louvain, Chemin des Étoiles, 8, B-1348 Louvain-la-Neuve, Belgium.

DOI [10.24435/materialscloud:c8-gy](https://doi.org/10.24435/materialscloud:c8-gy) [version v1]

Publication date: Nov 28, 2023

### Files

File name	Size	Description
mp_comparison.json.gz <a href="#">MDS</a>	49.0 KiB	A list of pymatgen 'ComputedEntry' containing the results of atomate2 'MPGGARelax' calculations to enable direct comparison with the Materials Project's convex hull (as of 20/11/2023).
data.csv <a href="#">MDS</a>	18.5 KiB	Computed ELF max and stability info for each structure
raw.tar.gz <a href="#">MDS</a>	328.2 MiB	Additional raw data, including bandstructures and phonon dispersion curves for every structure.
structures.tar.gz <a href="#">MDS</a>	40.4 KiB	CIFs of every structure considered, with the exact structures used for initial ELFCar calculations, as well as by the MP compatibility relaxation (in the corresponding sub-folders).
scripts.zip <a href="#">MDS</a>	2.9 KiB	A directory containing Python scripts used for the re-relaxation and stability calculations of the structures, alongside a requirements file with the dependencies required for repeating them, as well as the script required to create this archive from the raw data.
README.txt <a href="#">MDS</a>	1.0 KiB	README
optimade.yaml <a href="#">MDS</a> 	1.5 KiB	A config file for the MCloud/OPTIMADE integration that allows ingestion of the data into an OPTIMADE API

```
1 entries:
2   - entry_type: structures
3     entry_paths:
4       - file: structures.tar.gz
5         matches:
6           - structures/mp_gga/*.cif
7     property_paths:
8       - file: data.csv
9     property_definitions:
10      - name: elf_max
11        title: Maximum value of the ELF
12        description: The maximum value of the electron
13        unit: dimensionless
14        type: float
```

<https://optimadeclient.materialscloud.io>

Query a provider's database

Materials Cloud Archive  
Select a provider  
Alexandria  
Computational materials repository (CMR)  
Materials Cloud  
Materials Cloud Archive  
The Materials Project  
Material-Property-Descriptor Database  
Material Properties Open Database (MPOD)  
openBabel Database  
Open Materials Database  
2DMaterials  
AFLW  
Cryophysics Open Database  
Materials Platform for Data Science  
optmake  
novel materials discovery (NOMAD)  
The Open Quantum Materials Database (OQMD)  
Joint Automated Repository for Various Integrated Simulations (JARVIS)  
Theoretical Crystallography Open Database

Databases contributed by the community to the Materials Cloud Archive  
Z. Wang et al., Machine learning-accelerated discovery of A<sub>2</sub>BC<sub>2</sub> ternary electrides with diverse anionic electron densities, Materials Cloud Archive 2023.181 (2023) doi: [10.24435/materialscloud:c8-gy](https://doi.org/10.24435/materialscloud:c8-gy)

Results  
Showing 1-25 of 145 results  
Dy4Si4Y2 (id:structures.tar.gz/structures/mp\_gga)

Crystallographic Information File v1.0 [via ASE] ( cif ) Download  
Use In QE Input Generator

# Archive-OPTIMADE service

<https://archive.materialscloud.org/>

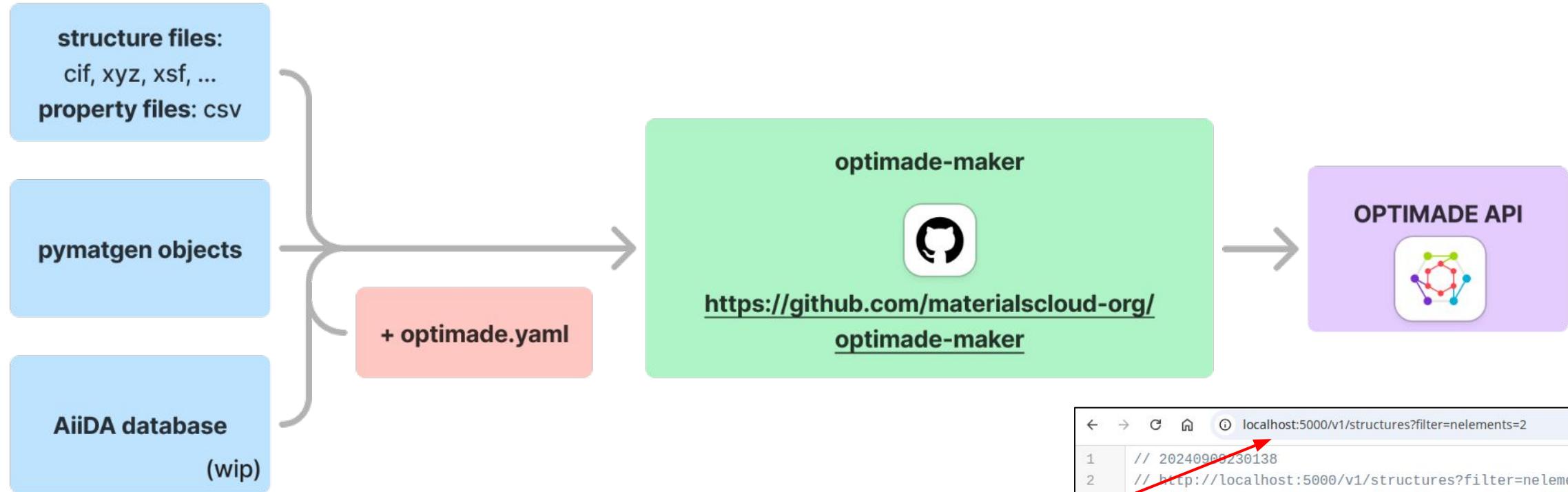
The screenshot shows the Materials Cloud Archive interface. At the top, there's a navigation bar with 'LEARN', 'WORK', 'DISCOVER', 'EXPLORE', 'ARCHIVE' (which is highlighted in red), 'More', 'Search', 'Upload a record', 'Log in', and 'Sign up'. Below the navigation, there are three main sections: 'SUBMISSION INSTRUCTIONS' (with a red arrow pointing to it), 'MODERATION POLICIES', and 'ARCHIVE POLICIES'. Under 'SUBMISSION INSTRUCTIONS', there's a heading 'Submission instructions' and a section titled 'To submit a record to the Materials Cloud Archive:' with a bulleted list. There are also links for 'Guidelines to fill in the upload form', 'Submission and moderation of records', and 'Update of published records'. A section titled 'Integration with external applications' follows, with subsections for 'hemiscope' (with a red arrow pointing to it) and 'OPTIMADE'. The 'OPTIMADE' section contains text about the OPTIMADE API and how to create an entry. At the bottom, there's a note about the full list of archive entries hosting an OPTIMADE API.

<https://optimade.materialscloud.org/>

The screenshot shows the Materials Cloud Archive OPTIMADE server interface. It features a header 'Materials Cloud Archive OPTIMADE server' with a logo of a molecular structure. Below the header, there are two sections: 'Index meta-database:' with links to '/archive/index/v1/info' and '/archive/index/v1/links', and 'Available databases:' with a table. The table has columns for 'Date', 'Archive entry', and 'OPTIMADE endpoint'. It lists three entries:

Date	Archive entry	OPTIMADE endpoint
2024.08.05	V. Trinquet et al., <i>Optical materials discovery and design with federated databases and machine learning</i> , Materials Cloud Archive <b>2024.114</b> (2024), doi: <a href="https://doi.org/10.24435/materialscloud:5p-vq">10.24435/materialscloud:5p-vq</a>	<a href="https://archive/5p-vq">/archive/5p-vq</a>
2024.04.26	L. Kahle et al., <i>High-throughput computational screening for solid-state Li-ion conductors</i> , Materials Cloud Archive <b>2024.65</b> (2024), doi: <a href="https://doi.org/10.24435/materialscloud:vg-ya">10.24435/materialscloud:vg-ya</a> [version 2]	<a href="https://archive/vg-ya">/archive/vg-ya</a>
2023.11.28	Z. Wang et al., <i>Machine learning-accelerated discovery of <math>A_2BC_2</math> ternary electrides with diverse anionic electron densities</i> , Materials Cloud Archive <b>2023.181</b> (2023), doi: <a href="https://doi.org/10.24435/materialscloud:c8-gy">10.24435/materialscloud:c8-gy</a>	<a href="https://archive/c8-gy">/archive/c8-gy</a>

# optimade-maker



```
> ls
data.csv optimade.yaml structures.zip
> optimake serve
2024-09-09 22:50:24 - optimake - INFO - optimade.jsonl doesn't exist. Converting archive.
Parsing structures files: 100%|██████████| 272 [00:00<00:00, 77.36
Constructing OPTIMADE structures entries: 2it [00:00, 731.29it/s]
Parsing properties for structures entries: 100%|██████████| 1/1 [00:00<00:00, 87.14
INFO: Started server process [31378]
INFO: Waiting for application startup.
INFO: Application startup complete.
INFO: Uvicorn running on http://0.0.0.0:5000 (Press CTRL+C to quit)
```

```
localhost:5000/v1/structures?filter=nelements=2
1 // 20240909230138
2 // http://localhost:5000/v1/structures?filter=nelements=2
3
4 {
5   "data": [
6     {
7       "id": "set2/102",
8       "type": "structures",
9       "attributes": {
10         "immutable_id": "structures.zip/cifs/set2/102.cif",
11         "last_modified": "2024-09-09T22:50:24.651087",
12         "elements": [
13           "C",
14           "Sr"
15         ],
16         "nelements": 2,
```

# Acknowledgements

People involved in AiiDA, AiiDALab, and Materials Cloud



Giovanni  
Pizzi  
(PSI)

Marnik  
Bercx  
(PSI)

Miki  
Bonacci  
(PSI)

Edan  
Bainglass  
(PSI)

Xing  
Wang  
(PSI)

Jusong  
Yu  
(PSI)

Ali  
Khosravi  
(PSI)

Alexander  
Goscinski  
(PSI)

Michail  
Minotakis  
(PSI)

Yuhao Jiang  
(PSI)

Timo  
Reents  
(PSI)

Nataliya  
Paulish  
(PSI)



Nicola  
Marzari  
(EPFL, PSI)



Sebastiaan  
P. Huber  
(Microsoft)



Chris  
Sewell  
(EPFL)



Francisco F.  
Ramirez  
(EPFL)



Kristjan  
Eimre  
(EPFL)



Daniel  
Hollas  
(U. Bristol)



Leopold  
Talirz  
(Schott)



Aliaksandr  
Yakutovich  
(Empa)



Carlo  
Pignedoli  
(Empa)



Valeria  
Granata  
(EPFL)



Joost  
VandeVondele  
(ETHZ,CSCS)



Thomas  
Schultheiss  
(ETHZ,CSCS)



swissuniversities



Materials Science and Technology



PARTNERSHIP FOR ADVANCED  
COMPUTING IN EUROPE

# Summary

---

- Recent AiiDA improvements:
  - verdi presto - easy installation
  - aiida-shell - run executables without AiiDA plugins
  - WorkGraphs - easy combining of AiiDA processes
- Materials Cloud 3D crystals database
  - 'PBEsol-v1' subdatabase
  - Improved web interface
- Materials Cloud Archive - OPTIMADE integration
  - automatic OPTIMADE APIs for contributed raw data
  - optimade-maker

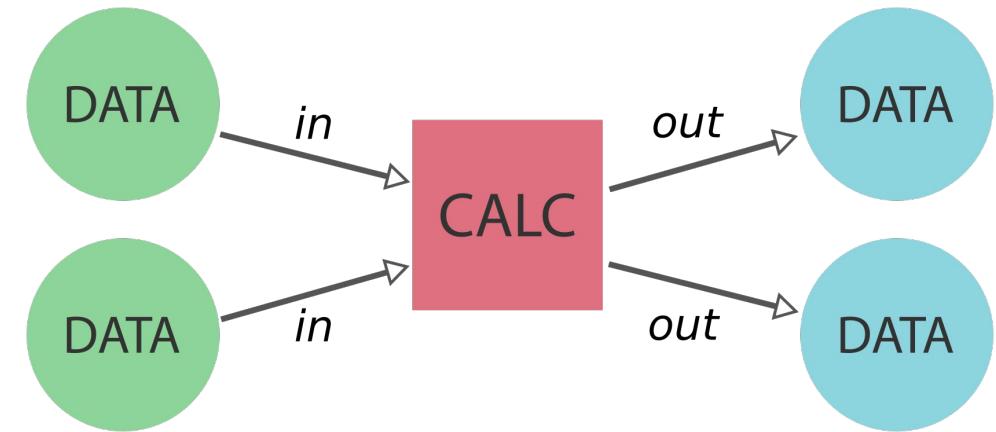


# Data provenance

---

Simple recipe - store:

- data transformations/calculations
- inputs
- outputs
- inter-connections



# Data provenance

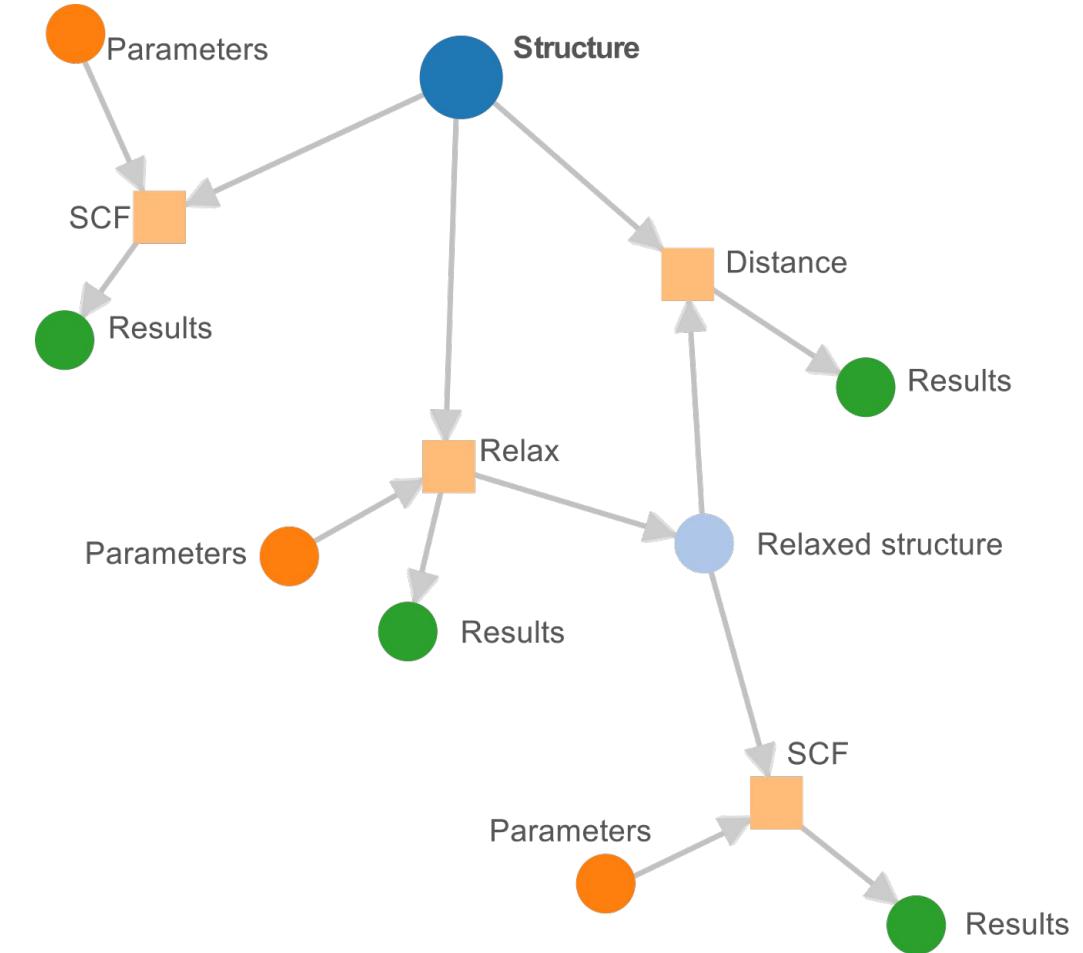
---

Simple recipe - store:

- data transformations/calculations
- inputs
- outputs
- inter-connections

Provenance graphs

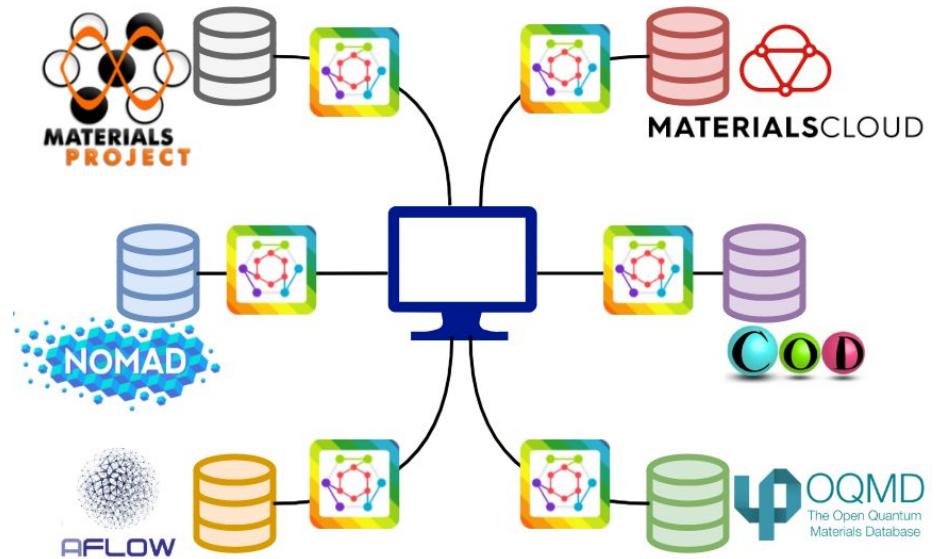
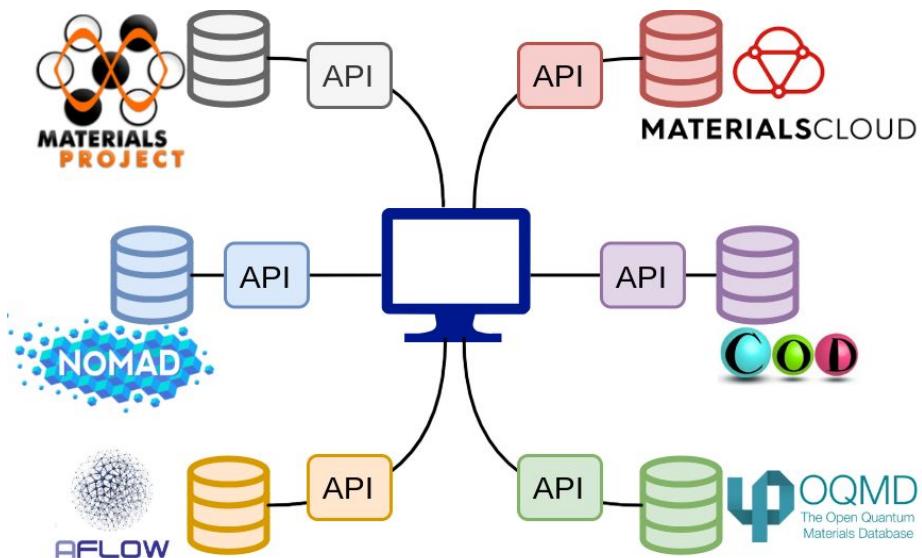
- when data gets reused, a directed graph is created
- that quickly grows in complexity



# OPTIMADE



Consortium to make materials databases  
interoperable via a **common REST API**, 15+ partners



# OPTIMADE



# OPTIMADE API

base URL specific to the DB

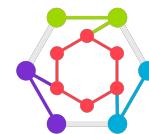
# universal OPTIMADE query

[https://aiida.materialscloud.org/mc3d/optimade/v1/structures?filter=elements HAS ALL "C", "Si"](https://aiida.materialscloud.org/mc3d/optimade/v1/structures?filter=elements HAS ALL \)

- supports queries for/based on
  - chemical composition;
  - custom properties;
  - references;
  - & more, with new functionality constantly being added

```
    "data": [
      {
        "id": "417",
        "type": "structures",
        "links": null,
        "meta": null,
        "attributes": {
          "immutable_id": "6dd18556-a5e6-442d-a0f2-da0ae5beb716",
          "last_modified": "2022-07-01T21:18:58Z",
          "elements": [
            "C",
            "Si"
          ],
          "nelements": 2,
          "elements_ratios": [ ... ], // 2 items
          "chemical_formula_descriptive": "C9Si9",
          "chemical_formula_reduced": "CSi",
          "chemical_formula_hill": "C9Si9",
          "chemical_formula_anonymous": "AB",
          "is_stable": true
        }
      }
    ]
  }
}
```

# OPTIMADE



## OPTIMADE Clients

<https://optimadeclient.materialscloud.io/>

The screenshot shows the OPTIMADE client interface. On the left, there's a sidebar with a logo, version information (v1.1.0, Client version 2022.9.19), and links for GitHub and the official web page. Below this is a "Query a provider's database" section with a dropdown menu showing various providers like CMR, Materials Cloud, Materials Project, etc. The main area has a search bar with "e.g., (H<sub>2</sub>O)<sub>2</sub> Na" and a "Basic" filter tab selected. A results table shows 1-20 of 3352 results for "Co<sub>8</sub>Ge<sub>12</sub>Li<sub>12</sub>O<sub>48</sub> (id=mp-1013807)". To the right is a 3D visualization of the crystal structure.

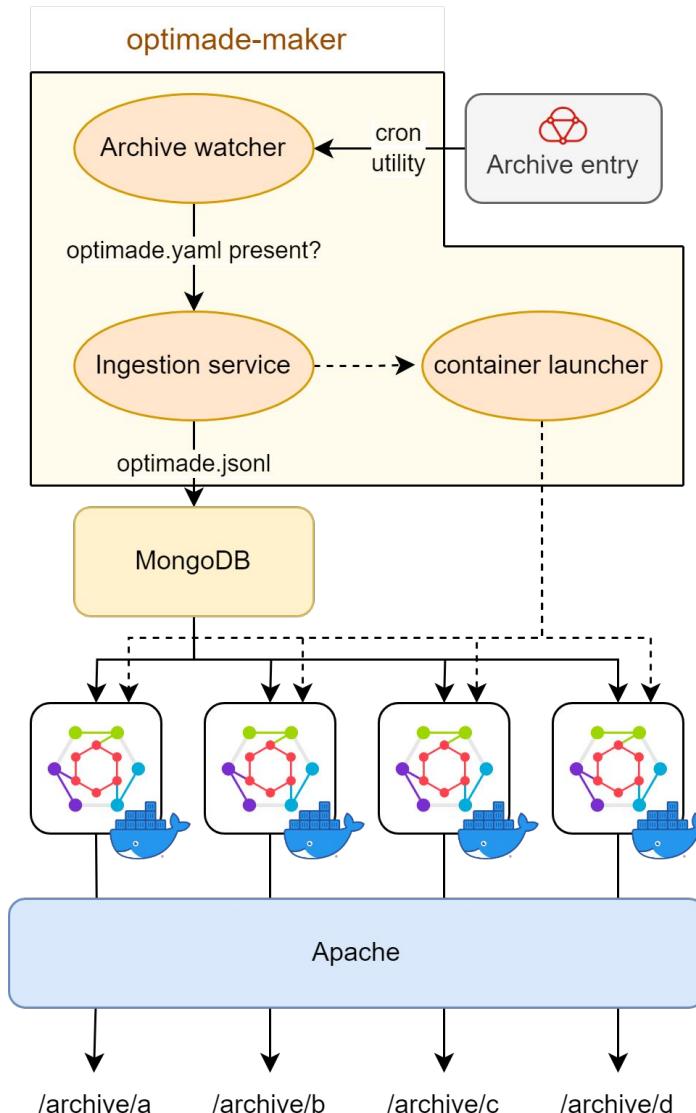
<https://www.materialscloud.org/work/tools/qeinputgenerator>

The screenshot shows the Quantum ESPRESSO input generator. It features a 3D visualization of a crystal structure with atoms represented by spheres and bonds by lines. To the right is a code editor window displaying Quantum ESPRESSO input code for a supercell. The code includes parameters like K\_POINTS, CELL\_PARAMETERS, and various atomic positions. Below the code editor is a "Generate the PWSCF input" button.

# Archive-OPTIMADE service

## Implementation

- Archive watcher detects optimade.yaml
- Pipeline to populate a MongoDB
- Docker containers for each API
  - OPTIMADE Python tools
- Apache as reverse proxy.
- Tools open source: optimade-maker
  - <https://github.com/materialscloud-org/optimade-maker>



# Materials Cloud OPTIMADE servers

<https://www.materialscloud.org/optimade>

The screenshot shows the Materials Cloud OPTIMADE APIs homepage. At the top, there is a navigation bar with icons for LEARN, WORK, DISCOVER, EXPLORE, ARCHIVE, and a More dropdown. Below the navigation bar, the breadcrumb navigation shows Home > OPTIMADE APIs. The main title is "Materials Cloud OPTIMADE APIs". Below the title, there are two icons: one for MAIN databases (a molecular structure) and one for ARCHIVE databases (a cloud icon). The MAIN section contains text about the provider name (mcloud), databases corresponding to Discover and Explore sections, and index meta-databases. The ARCHIVE section contains text about the provider name (mcloudarchive), databases served from the archive, and index meta-databases. A button at the bottom right of the ARCHIVE section says "Go to a table of the entries!".

LEARN WORK DISCOVER EXPLORE ARCHIVE More ▾

Home > OPTIMADE APIs

## Materials Cloud OPTIMADE APIs

**MAIN**

Provider name: mcloud  
Databases corresponding to the Materials Cloud Discover and Explore sections.  
Index meta-database:  
</optimade/main/v1/info>  
</optimade/main/v1/links>

**ARCHIVE**

Provider name: mcloudarchive  
Databases served automatically from the entries of the Materials Cloud Archive.  
Index meta-database:  
</optimade/archive/v1/info>  
</optimade/archive/v1/links>

Go to a table of the entries!