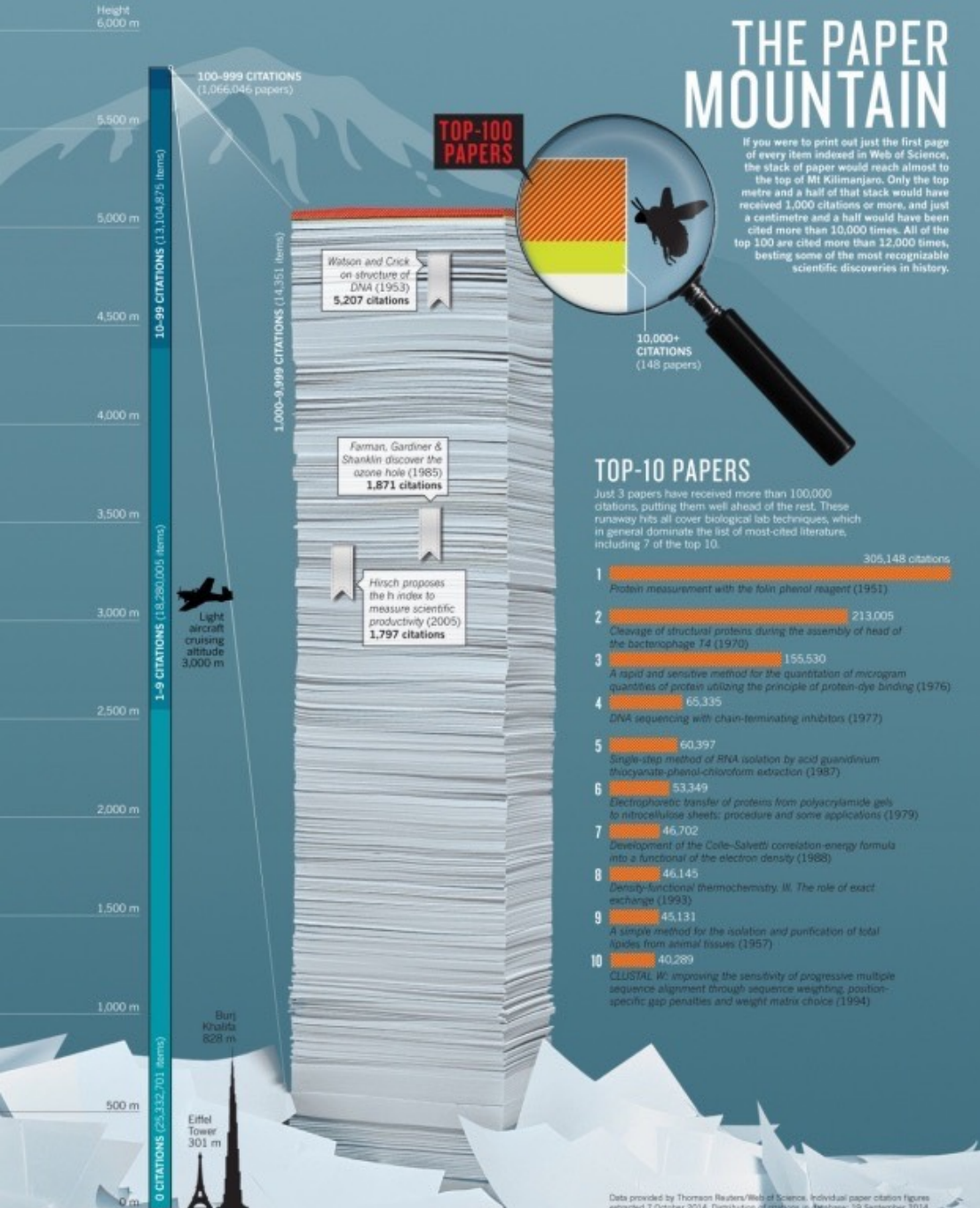




Managing computational materials science in the high-throughput era

The **ADES** model and the **AiiDA** infrastructure

Giovanni Pizzi (EPFL, Switzerland)



Nature, November 2014

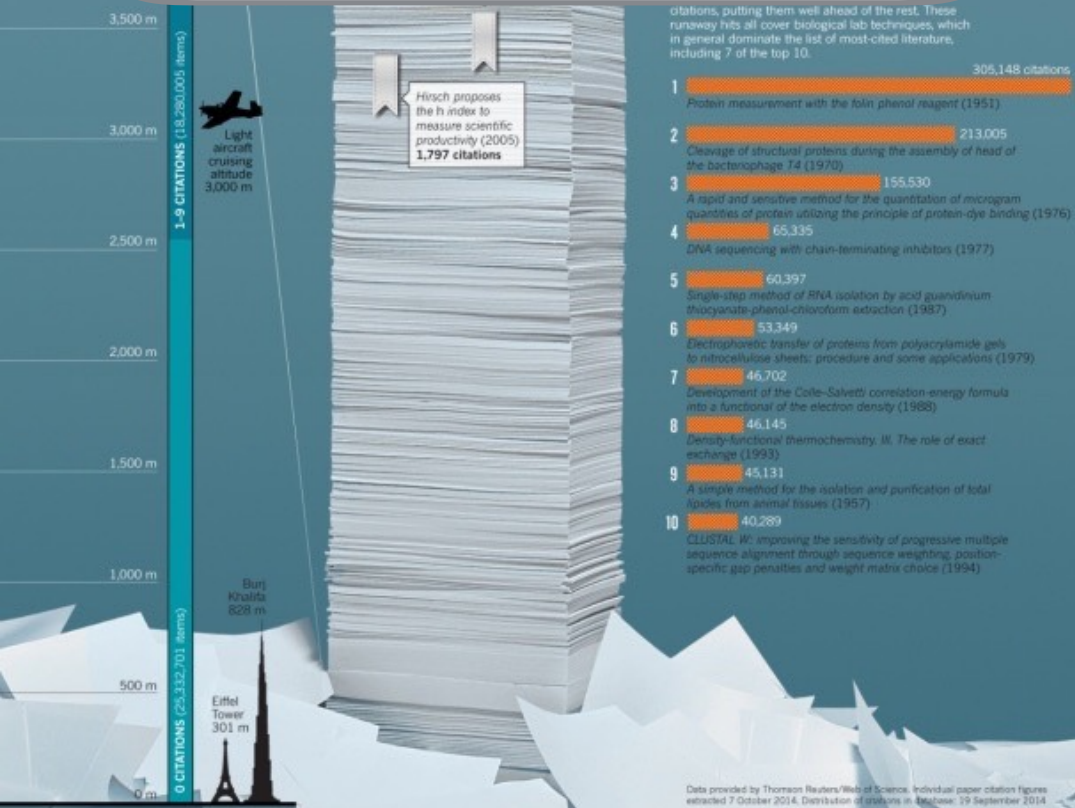
12 papers on DFT in the top-100 most cited papers in the entire scientific literature





first page of Science, almost to the top of the page, and just have been All of the 100 times, applicable in history.

Accuracy and predictive power of quantum engines



itations, putting them well ahead of the rest. These runaway hits all cover biological lab techniques, which in general dominate the list of most-cited literature, including 7 of the top 10.

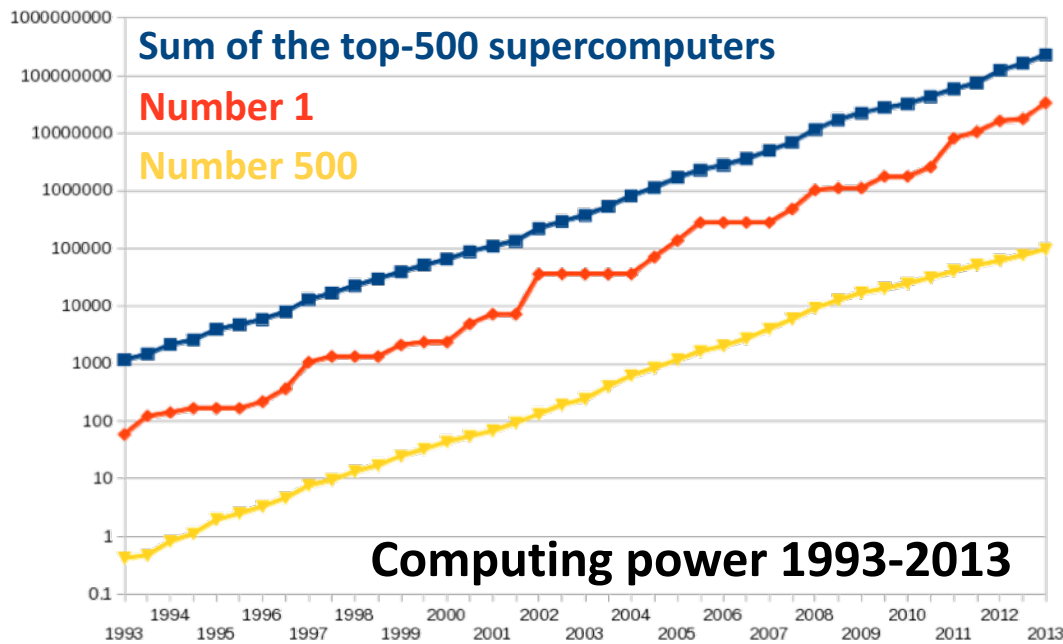
Hirsch proposes the h index to measure scientific productivity (2005)
1,797 citations

Data provided by Thomson Reuters/Web of Science. Individual paper citation figures extracted 7 October 2014. Distribution of citations in *Nature*: 19 September 2014.





Accuracy and predictive power of quantum engines



150,000x increase in the past 20 years

1 month (1993)

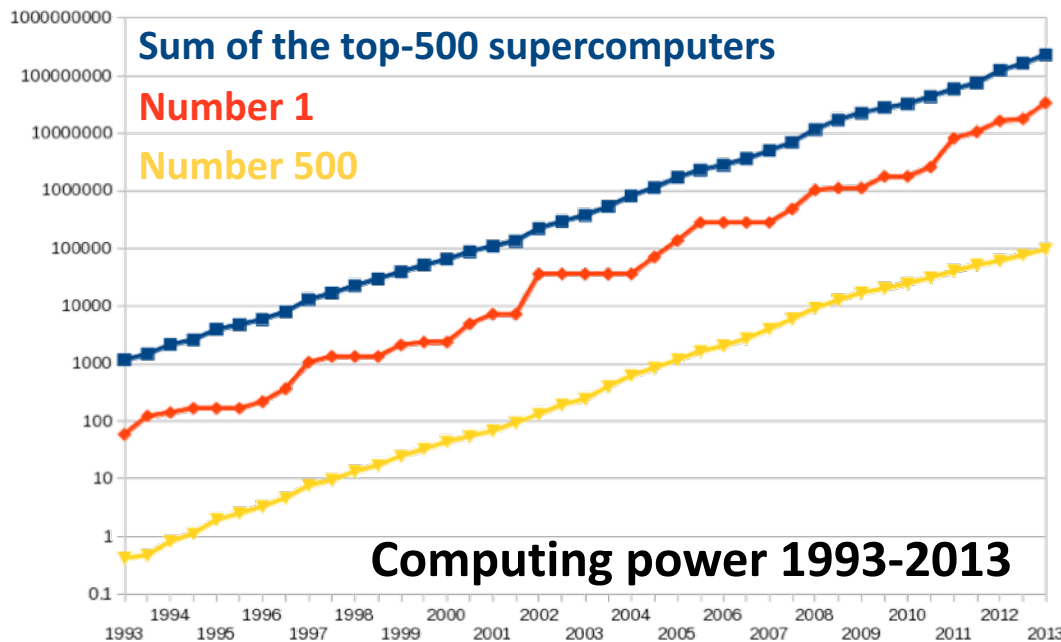


10 seconds (2015)





Accuracy and predictive power of quantum engines



150,000x increase in the past 20 years

1 month (1993)

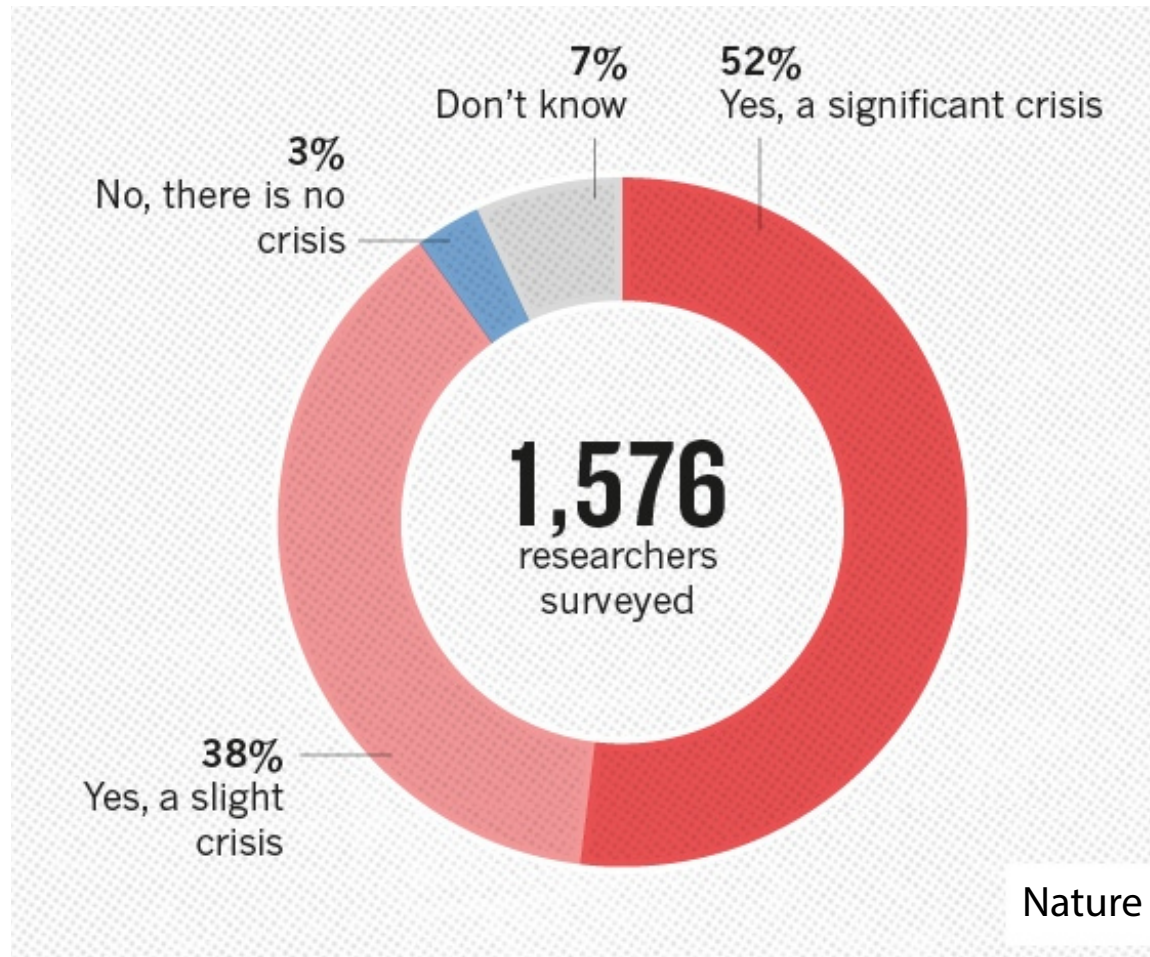


10 seconds (2015)

Result: materials design and discovery via
high-throughput computations

Reproducibility: a cornerstone of the scientific method

IS THERE A REPRODUCIBILITY CRISIS?



Reproducibility: a cornerstone of the scientific method

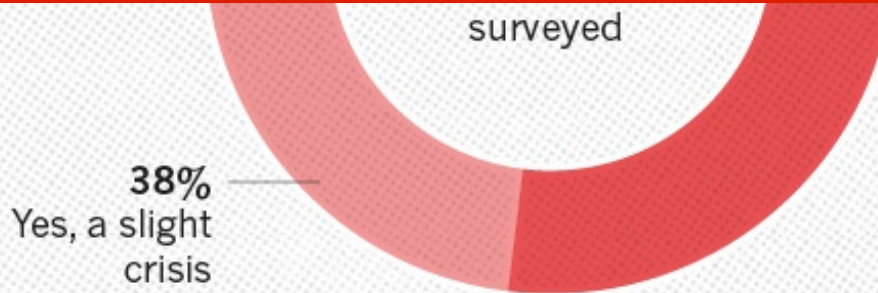
IS THERE A REPRODUCIBILITY CRISIS?

30% 7% 52%
Don't know Yes, a significant crisis

No excuses in **computational** science

We can and **must** be reproducible

Especially with the advent of **high-throughput**



Nature **533**, 452–454 (2016)



Data provenance and the importance of the 'how'

PROV·E·NANCE | 'präv-nən(t)s, 'prä-və-,nän(t)s

1. The origin or source of something
2. The history of ownership of a valued object or work of art or literature



Data provenance and the importance of the 'how'

PROV·E·NANCE | 'präv-nən(t)s, 'prä-və-,nän(t)s

1. The origin or source of something
2. The history of ownership of a valued object or work of art or literature

To guarantee reproducibility it is not enough to just store the

WHAT

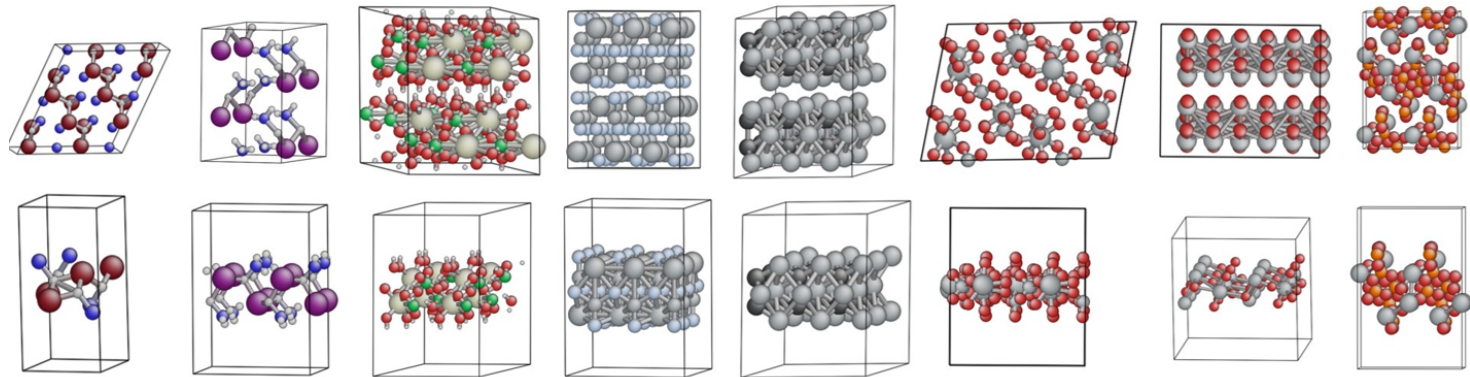
but it is crucial to also store the

HOW



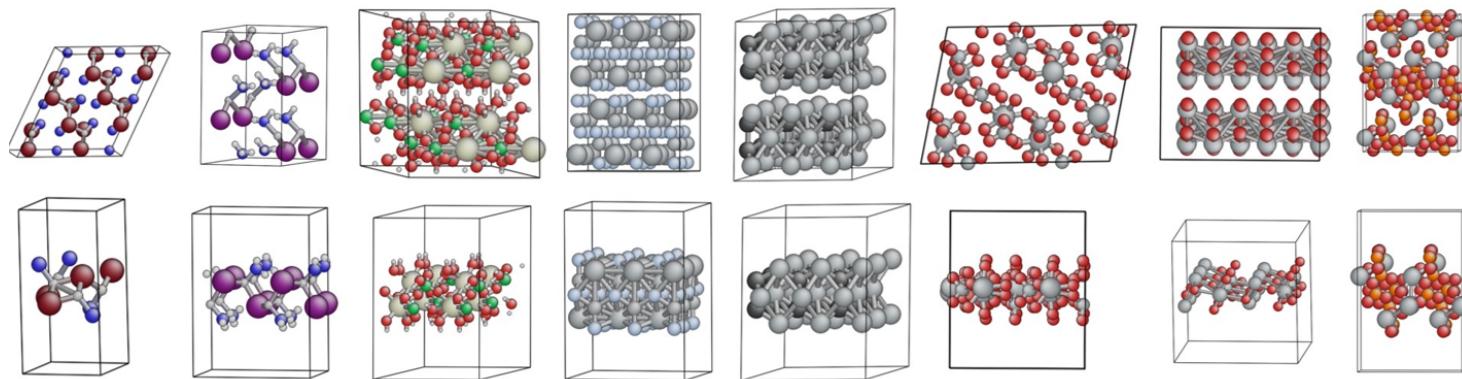
Importance of provenance: an high-throughput example

DISCOVERING NEW TWO-DIMENSIONAL MATERIALS



Importance of provenance: an high-throughput example

DISCOVERING NEW TWO-DIMENSIONAL MATERIALS



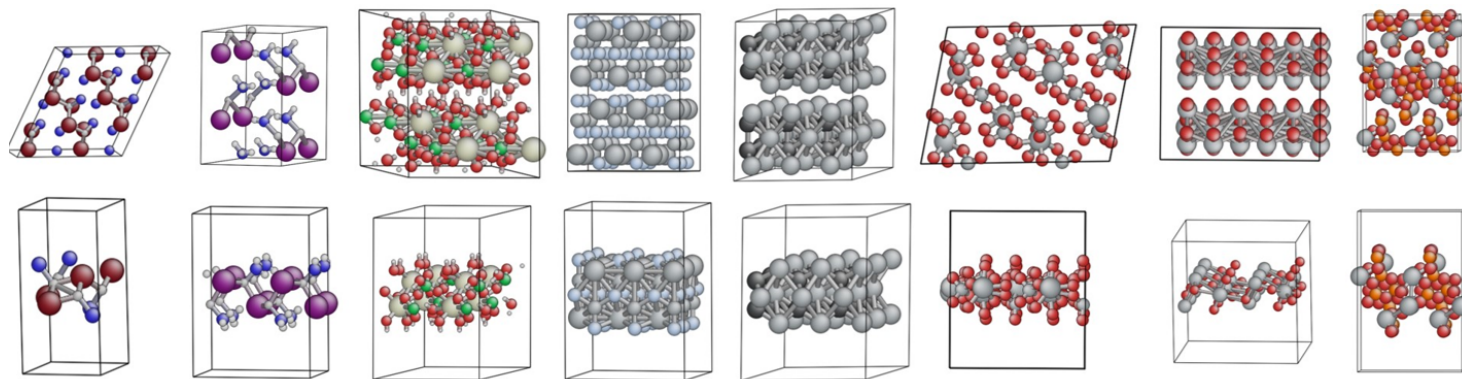
STARTING FROM ICSD/COD DATABASE:

- **108 423** unique 3D structures
- **5619** layered structures
- **>100 000** DFT calculations
- **>30 000** material properties
- **$>1 \cdot 10^9$** attributes



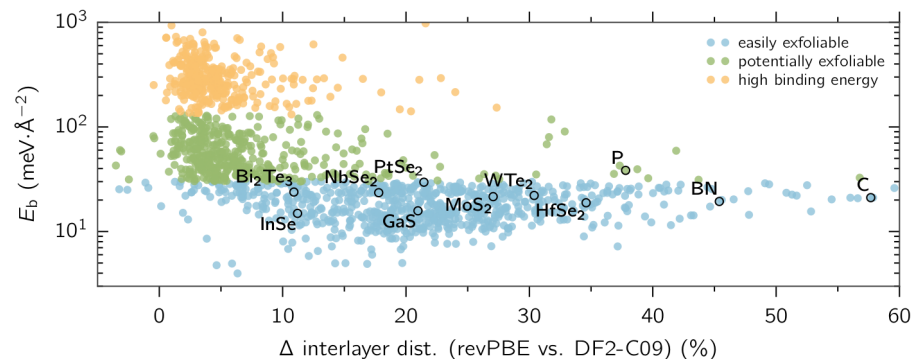
Importance of provenance: an high-throughput example

DISCOVERING NEW TWO-DIMENSIONAL MATERIALS



STARTING FROM ICSD/COD DATABASE:

- **108 423** unique 3D structures
- **5619** layered structures
- **>100 000** DFT calculations
- **>30 000** material properties
- **>1·10⁹** attributes

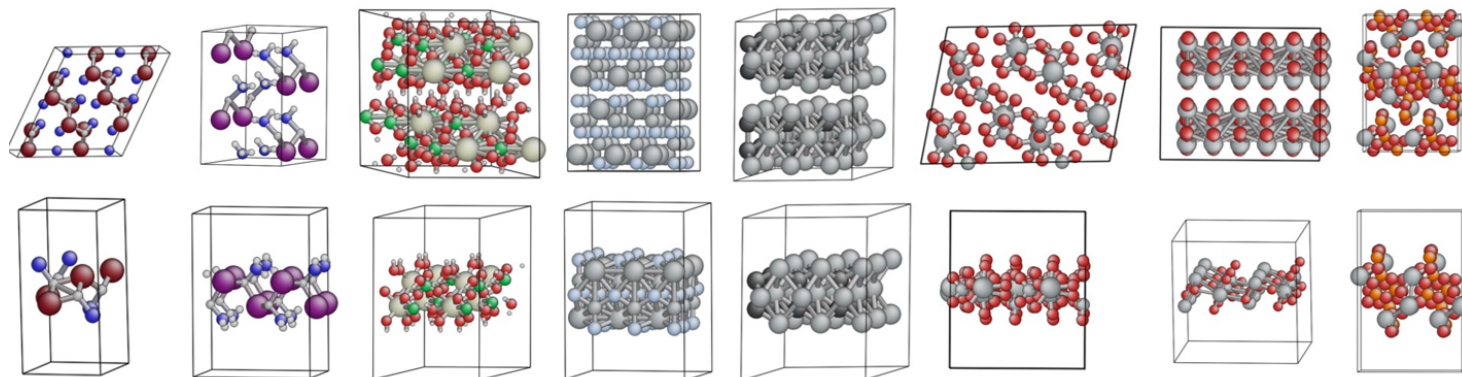


All data needs to be condensed in a few plots



Importance of provenance: an high-throughput example

DISCOVERING NEW TWO-DIMENSIONAL MATERIALS



STARTING FROM ICSD/COD DATABASE:

- **108 423** unique 3D structures
- **5619** layered structures
- **>100 000** DFT calculations
- **>30 000** material properties
- **>1·10⁹** attributes

III. METHODS

A. Reproducibility and provenance

It is an often repeated tenet that results of scientific research must be reproducible. This objective is, however, challenging, especially in high-throughput research, due to the large number of simulations involved and the complex sequence of logical steps needed in the study. To ensure reproducibility we use AiiDA^[30] as a materials' informatics infrastructure to implement the ADES model of automation, data, environment, and sharing, as dis-

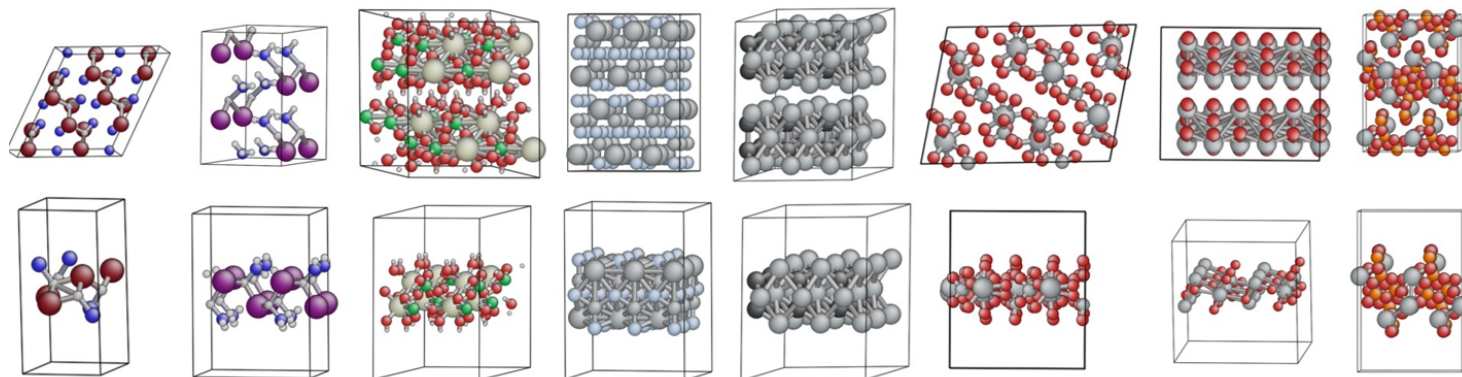
As already mentioned in the main text, in this step we consider only structures coming from experimental measurements, while the source databases are partially populated by purely theoretical structures. This is done using the flags set on the database entries by their curators. We also implemented some heuristics to detect clearly wrong CIF files. For example, we discard structures where the chemical formula provided in the file is inconsistent with the elements in the unit cell. Regardless of all these efforts, it is possible that some incomplete or incorrect structures are still not filtered out from the original databases.

Impossible to describe every detail in the Methods



Importance of provenance: an high-throughput example

DISCOVERING NEW TWO-DIMENSIONAL MATERIALS



STARTING FROM ICSD/COD DATABASE:

- **108 423** unique 3D structures
- **5619** layered structures
- **>100 000** DFT calculations
- **>30 000** material properties
- **>1·10⁹** attributes

III. METHODS

A. Reproducibility and provenance

It is an often repeated tenet that results of scientific research must be reproducible. This objective is, however, challenging, especially in high-throughput research, due to the large number of simulations involved and the complex sequence of logical steps needed in the study. To ensure reproducibility we use AiiDA³⁰ as a materials' informatics infrastructure to implement the ADES model of automation, data, environment, and sharing, as dis-

As already mentioned in the main text, in this step we consider only structures coming from experimental measurements, while the source databases are partially populated by purely theoretical structures. This is done using the flags set on the database entries by their curators. We also implemented some heuristics to detect clearly wrong CIF files. For example, we discard structures where the chemical formula provided in the file is inconsistent with the elements in the unit cell. Regardless of all these efforts, it is possible that some incomplete or incorrect structures are still not filtered out from the original databases.

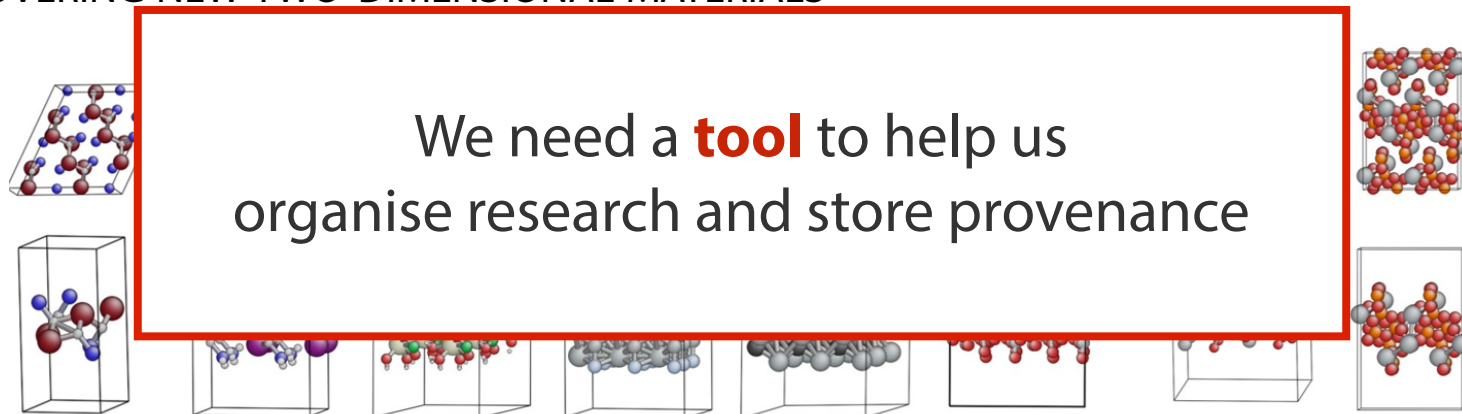
Impossible to describe every detail in the Methods

*For **peers**, reproducing data is challenging if not impossible
But even for **authors**, it is extremely difficult*



Importance of provenance: an high-throughput example

DISCOVERING NEW TWO-DIMENSIONAL MATERIALS



STARTING FROM ICSD/COD DATABASE:

- **108 423** unique 3D structures
- **5619** layered structures
- **>100 000** DFT calculations
- **>30 000** material properties
- **>1·10⁹** attributes

III. METHODS

A. Reproducibility and provenance

It is an often repeated tenet that results of scientific research must be reproducible. This objective is, however, challenging, especially in high-throughput research, due to the large number of simulations involved and the complex sequence of logical steps needed in the study. To ensure reproducibility we use AiiDA³⁰ as a materials' informatics infrastructure to implement the ADES model of automation, data, environment, and sharing, as dis-

As already mentioned in the main text, in this step we consider only structures coming from experimental measurements, while the source databases are partially populated by purely theoretical structures. This is done using the flags set on the database entries by their curators. We also implemented some heuristics to detect clearly wrong CIF files. For example, we discard structures where the chemical formula provided in the file is inconsistent with the elements in the unit cell. Regardless of all these efforts, it is possible that some incomplete or incorrect structures are still not filtered out from the original databases.

Impossible to describe every detail in the Methods

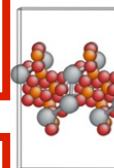
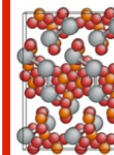
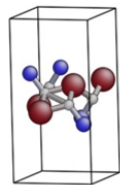
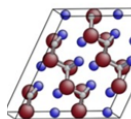
*For **peers**, reproducing data is challenging if not impossible
But even for **authors**, it is extremely difficult*



Importance of provenance: an high-throughput example

DISCOVERING NEW TWO-DIMENSIONAL MATERIALS

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



STARTING FROM

- **108 423** unit cells
- **5619** layered structures
- **>100 000** DFT calculations
- **>30 000** materials
- **>1·10⁹** attributes



Automated Interactive Infrastructure and
Database for Computational Science

- Computational science **platform**
- for **high-throughput** calculations
- with **automatic data provenance**

mentioned in the main text, in this step we filter out structures coming from experimental sources. The source databases are partially filtered by theoretical structures. This is done by adding the source database to the database entries by their complemented some heuristics to detect errors. For example, we discard structural formula provided in the file is not a valid chemical formula. Elements in the unit cell. Regarding, it is possible that some incomplete entries are still not filtered out from the

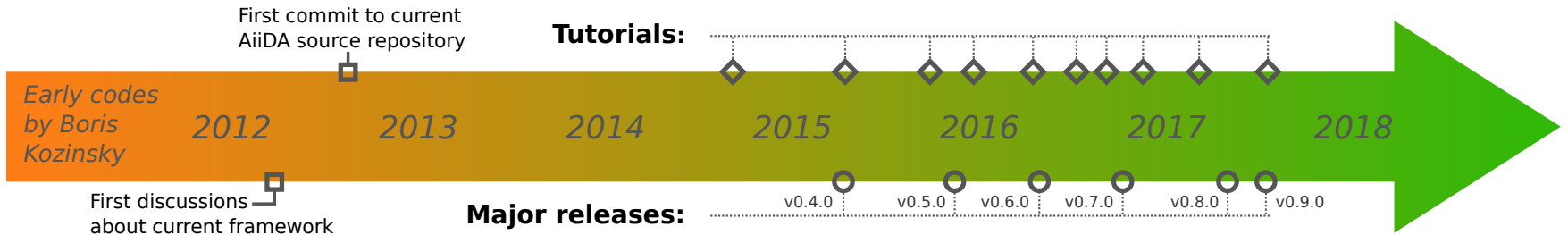
in the Methods





Automated Interactive Infrastructure and Database for Computational Science

The history of the code



https://github.com/a iidateam/a iida_core

<http://a iida.net>

The screenshot shows the GitHub repository for AiiDA. It includes the repository name 'a iidateam/a iida_core', statistics (5,996 commits, 6 branches, 17 releases, 26 contributors), and a list of files and folders such as .travis-data, a iida, bin, docs, examples, and utils.

The screenshot shows the AiiDA website homepage. It features the AiiDA logo, the tagline 'Automated Interactive Infrastructure and Database for Computational Science', and sections for 'Tutorials', 'Downloadable tutorial', 'Latest version', 'Older versions', 'Short video tutorials', and 'Events'. It also includes a 'MIT License' badge and the 'open source initiative' logo.

Such tool should help make computational research:

reproducible

Often not possible from the data
reported in papers

searchable

Find existing calculations,
reuse and data-mine results

reliable

Automated workflows to
verify results and
reduce the chance of errors

shareable

Community to share results,
cross-validate them, and boost
scientific discovery



Such tool should help make computational research:

reproducible

Often not possible from the data reported in papers

searchable

Find existing calculations, reuse and data-mine results

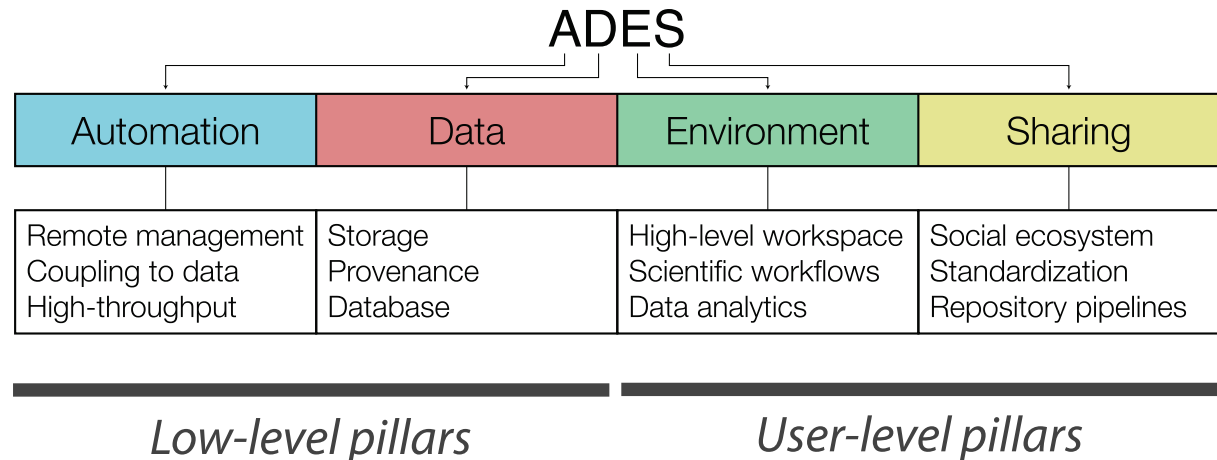
reliable

Automated workflows to verify results and reduce the chance of errors

shareable

Community to share results, cross-validate them, and boost scientific discovery

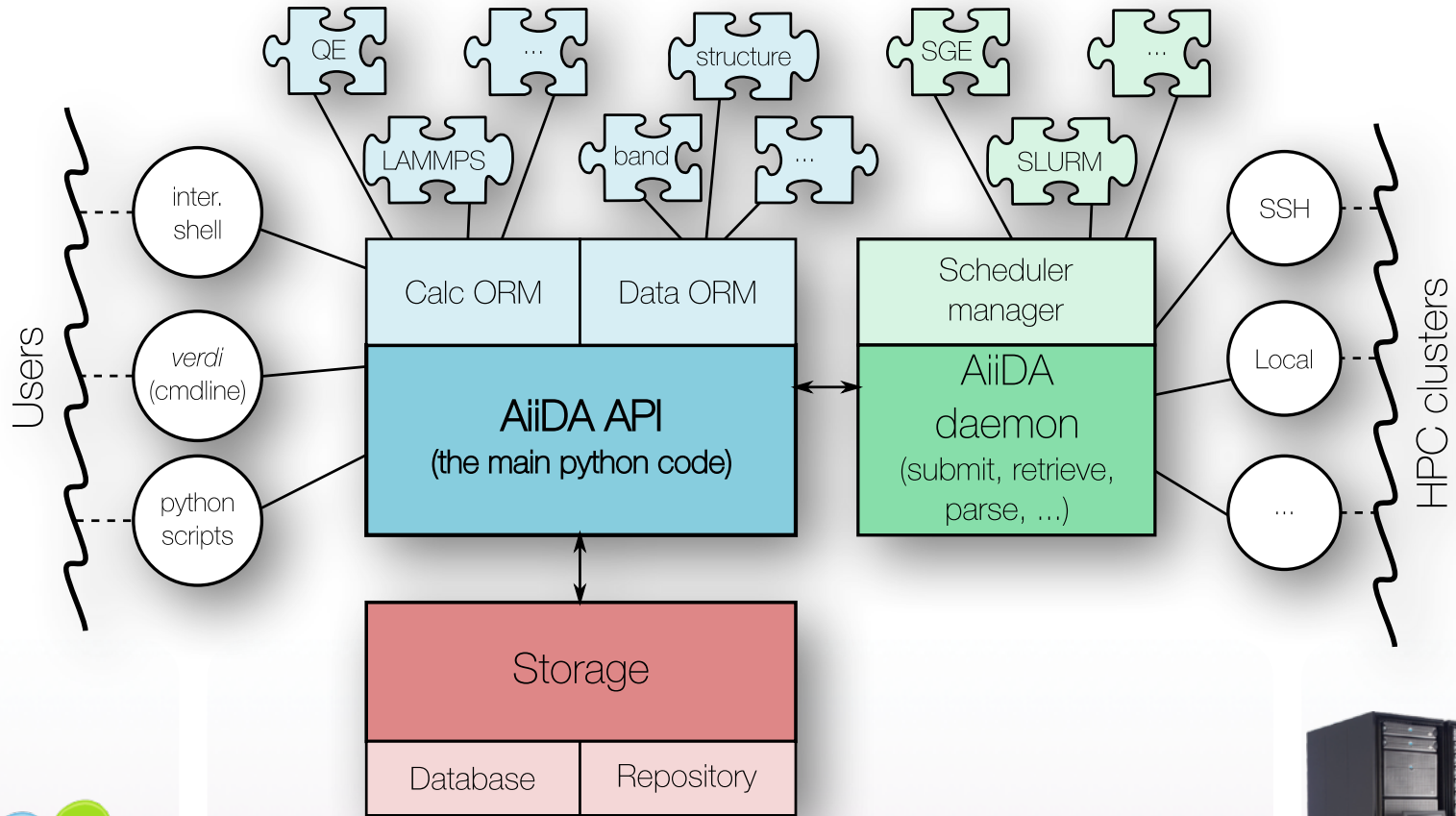
We have encoded these requirements in the four **ADES** pillars for a computational science infrastructure



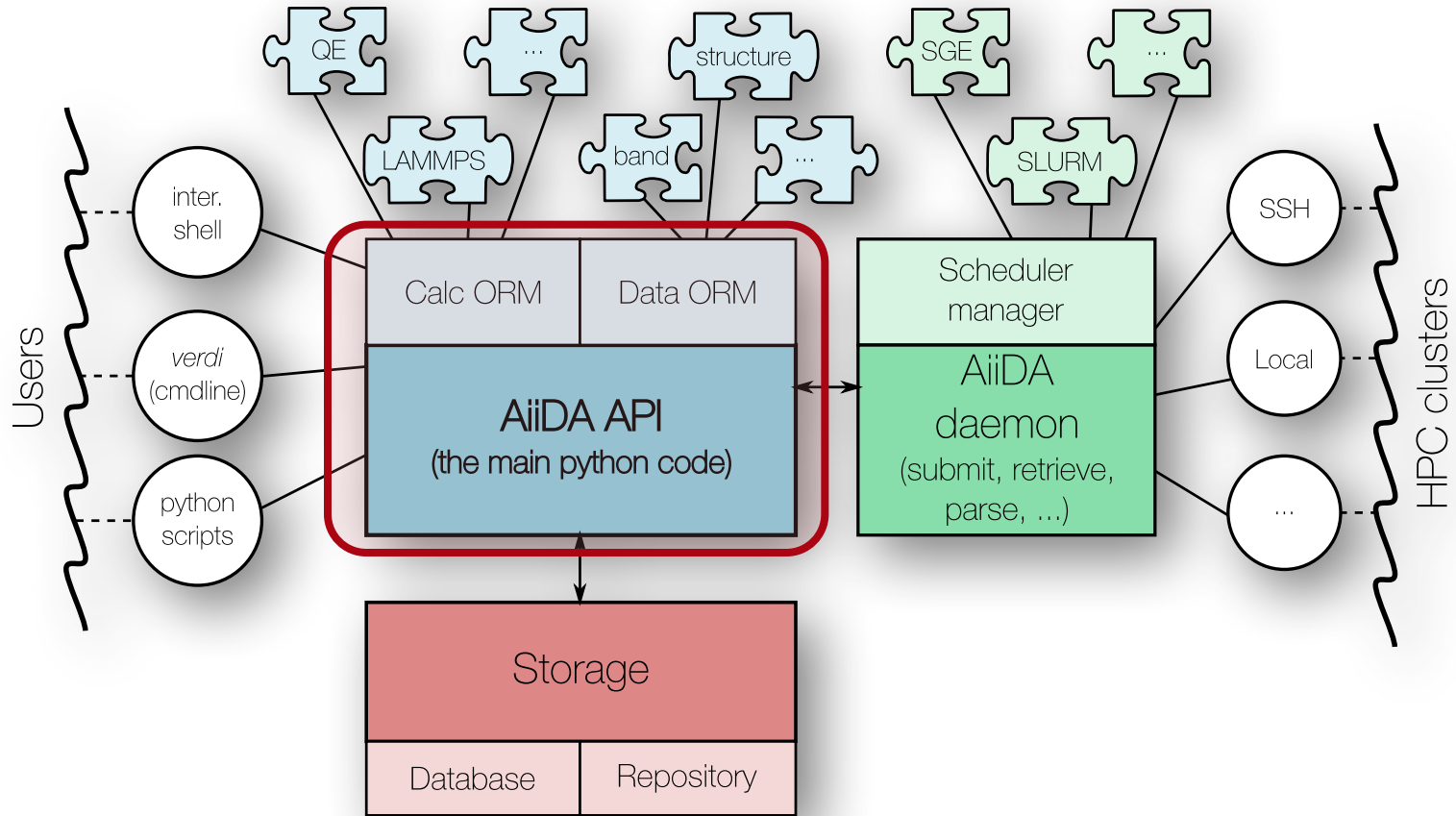
G. Pizzi et al., Comp. Mat. Sci. 111, 218-230 (2016)



What is AiiDA?



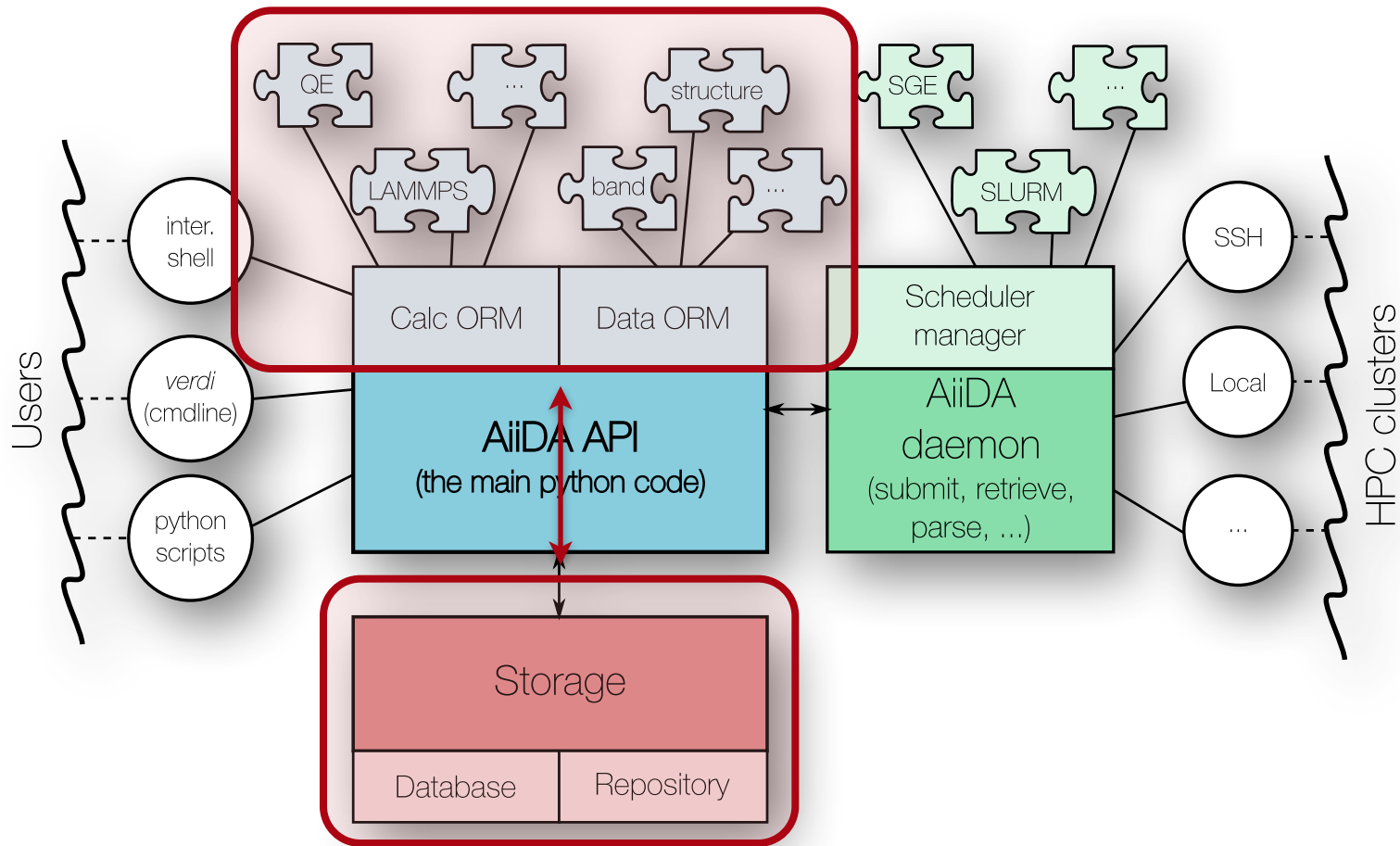
What is AiiDA?



1. The core of the code is the **AiiDA API**: the main python code and classes

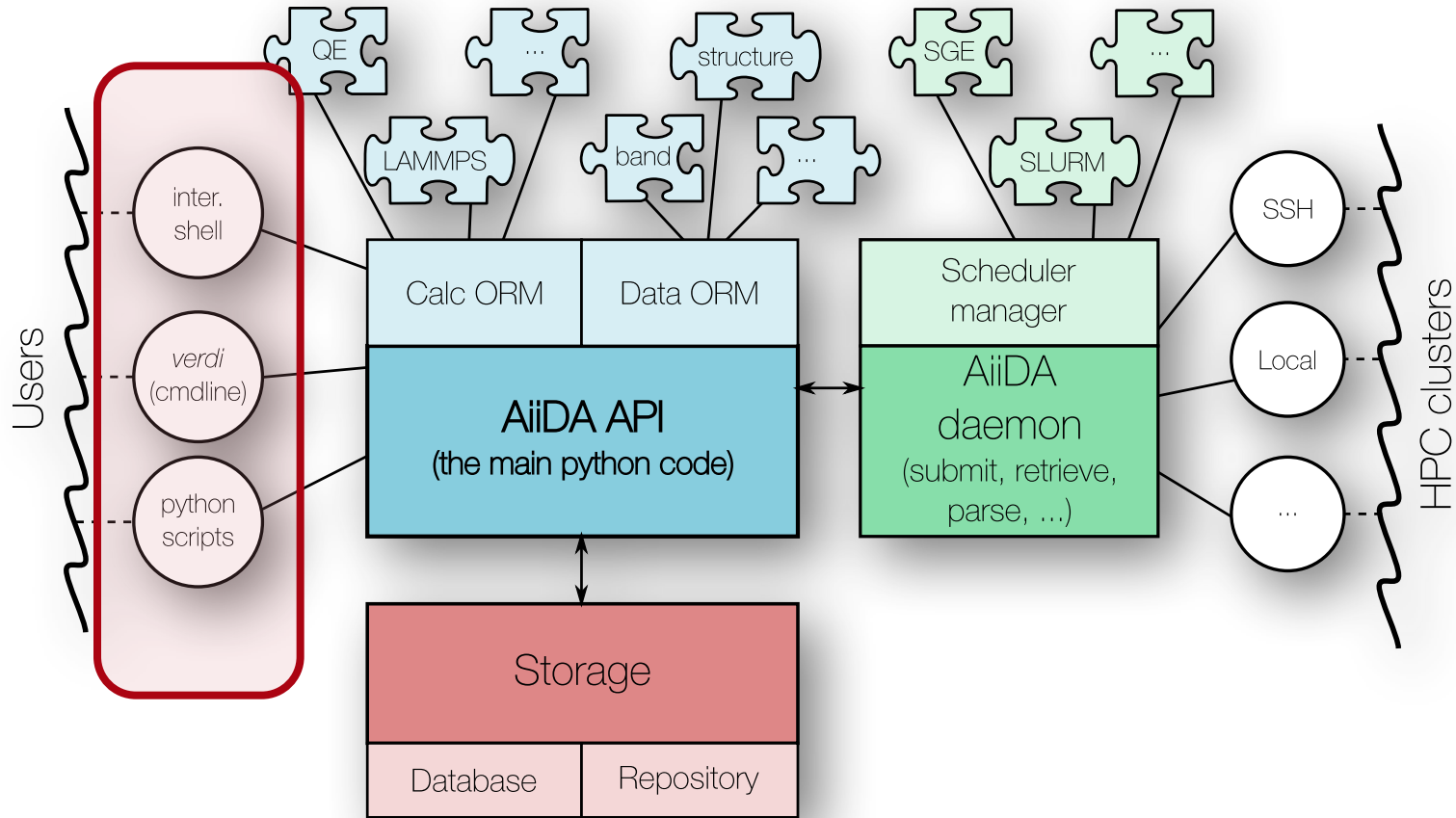


What is AiiDA?



2. The **AiiDA Object-Relational Mapper** (ORM): transparently store data, codes and calculations in a database, *transparent to the user*

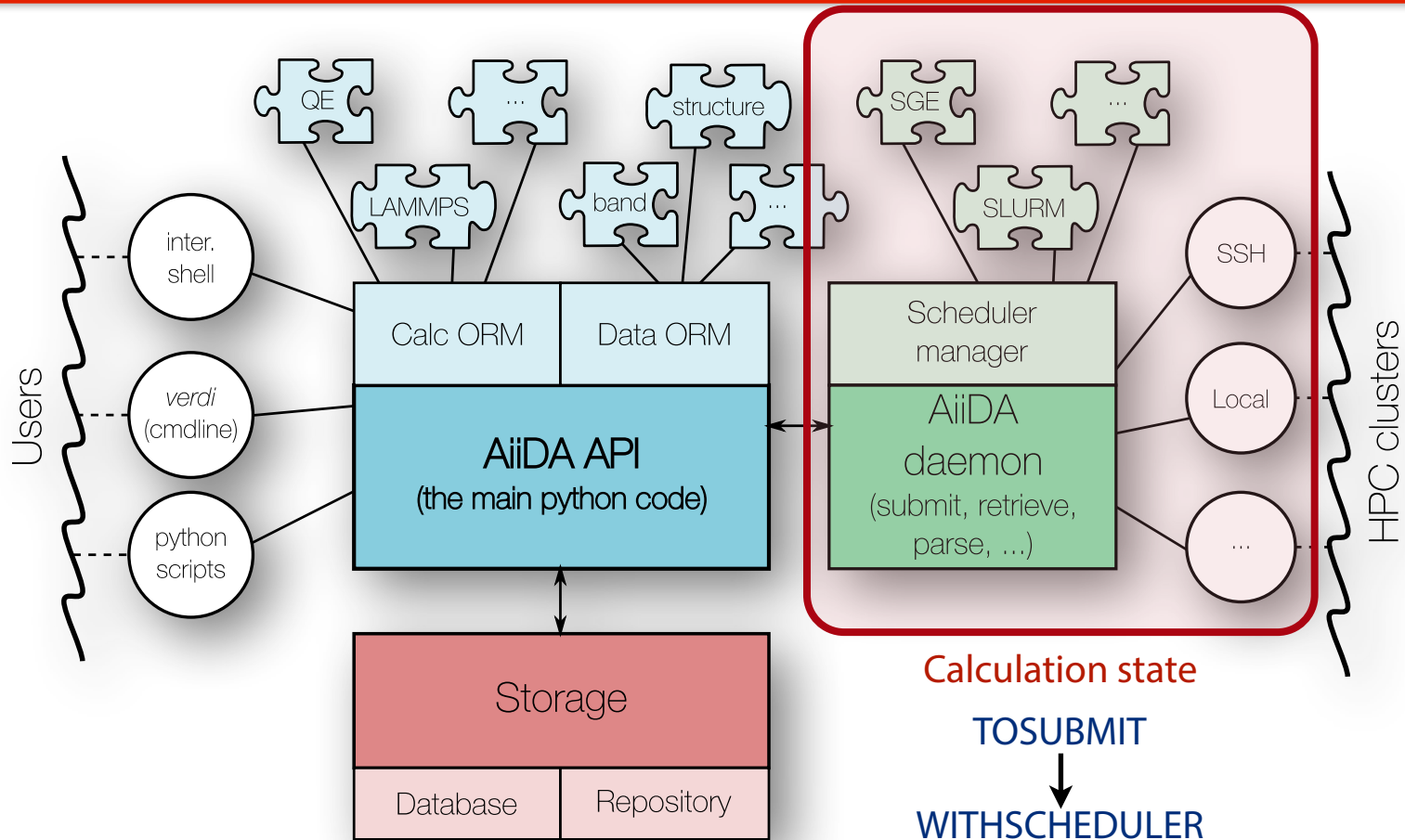
What is AiiDA?



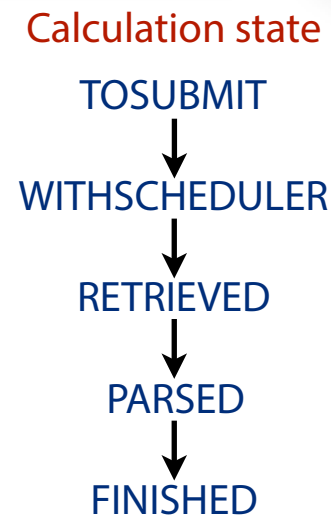
3. **User interaction** occurs via the command line tool **verdi**, the interactive shell or via Python scripts



What is AiiDA?

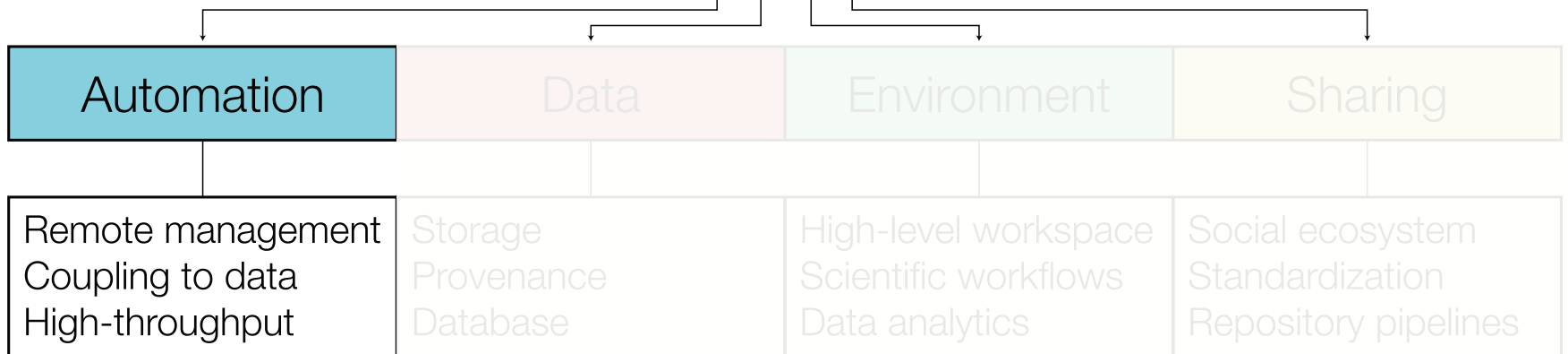


4. A **daemon** to manage interaction with remote computers without user intervention



Automation in AiiDA

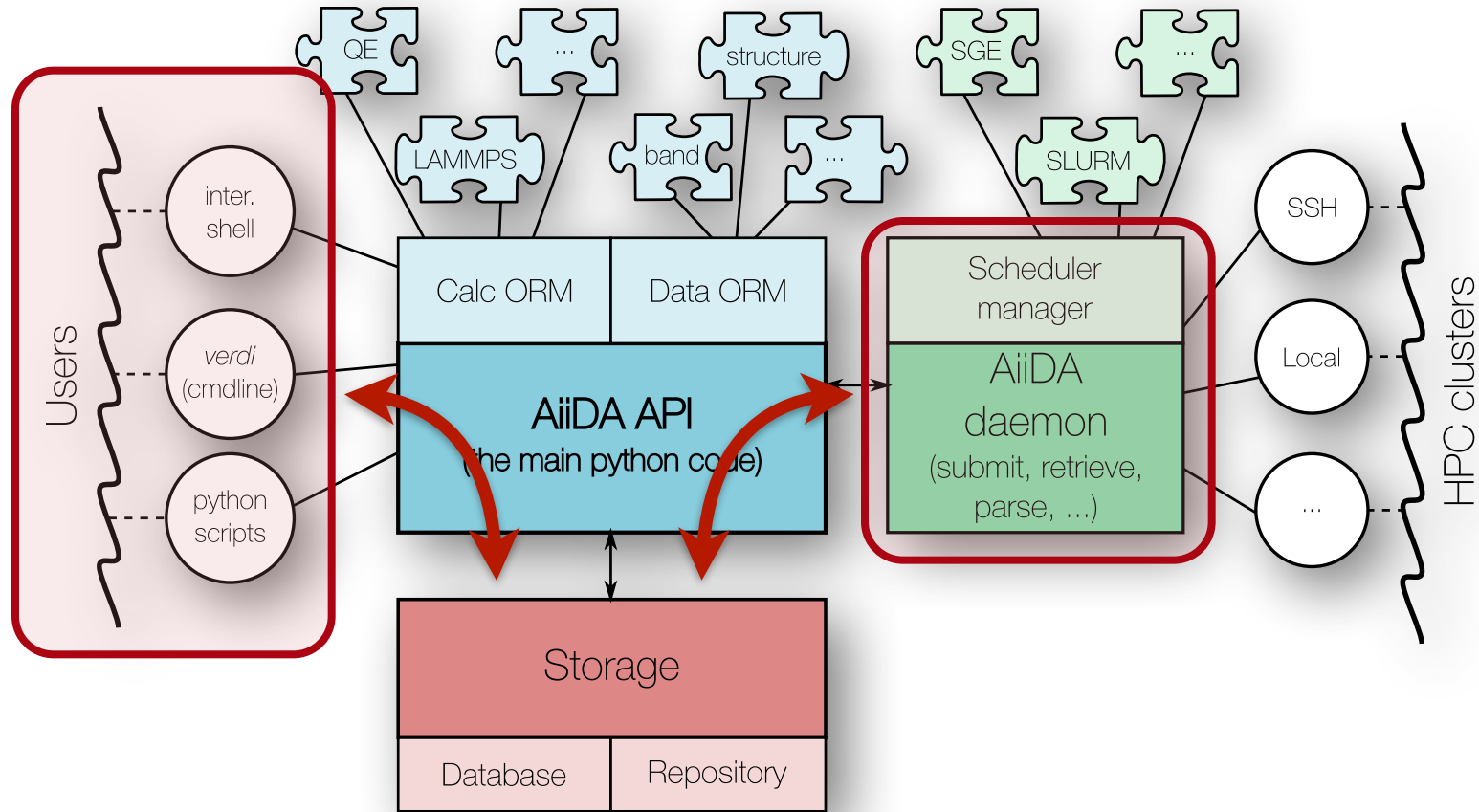
ADES



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



Automation: coupling to data



- **Coupling automation to data:**

- *uniformity* of the input data, usage of codes and computers
- *full reproducibility* of calculations (data is stored first)



Abstraction into APIs: a single calculation

```
Parameter = DataFactory('parameter')
Structure = DataFactory('structure')

code = get_code('quantumespresso-pw@mycluster')
JobCalc = code.new_process()
```

Choose code and computer

```
attrs = {
    'max_wallclock_seconds': 3600,
    'resources': {"num_machines": 2},
}
```

```
inp = {}
inp['structure'] = Structure(cif='silicon.cif')
```

Define all inputs

```
inp['parameters'] = Parameter({
    'CONTROL': {
        'calculation': 'scf',
        'restart_mode': 'from_scratch',
    },
    'SYSTEM': {
        'ecutwfc': 40.,
    }
})
```

```
f = submit(JobCalc, _attributes=attrs, **inp)
```

```
print f.job.pk
```

Take care of running the calculation
through the daemon



Abstraction into APIs: a single calculation

```
Parameter = DataFactory('parameter')
Structure = DataFactory('structure')

code = get_code('quantumespresso-pw@cluster2')
JobCalc = code.new_process()

attrs = {
    'max_wallclock_seconds': 3600,
    'resources': {"num_machines": 2},
}

inp = {}
inp['structure'] = Structure(cif='silicon.cif')

inp['parameters'] = Parameter({
    'CONTROL': {
        'calculation': 'scf',
        'restart_mode': 'from_scratch',
    },
    'SYSTEM': {
        'ecutwfc': 40.,
    }
})

f = submit(JobCalc, _attributes=attrs, **inp)

print f.job.pk
```

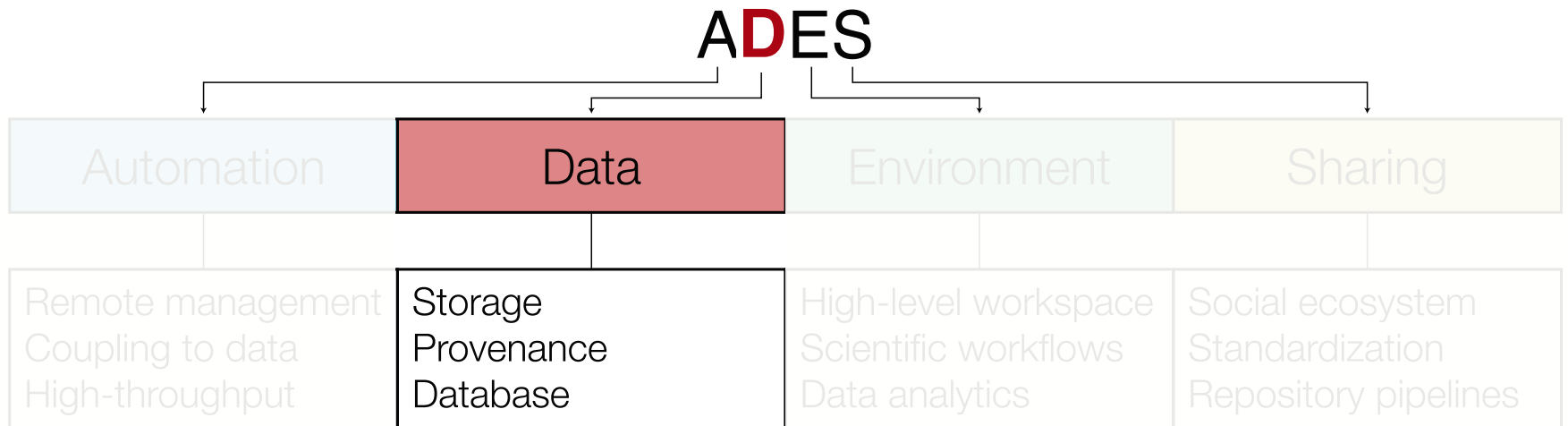
Just one line change to change computer (with different cluster, schedulers, version of codes, ...)

Data gets stored in the DB during submission

Take care of running the calculation through the daemon



Data in AiiDA



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



Storage and provenance

- *Calculated properties*: result of complex, connected calculations
- How do we store simulations **preserving the connected structure?**



Storage and provenance

- *Calculated properties*: result of complex, connected calculations
- How do we store simulations **preserving the connected structure?**

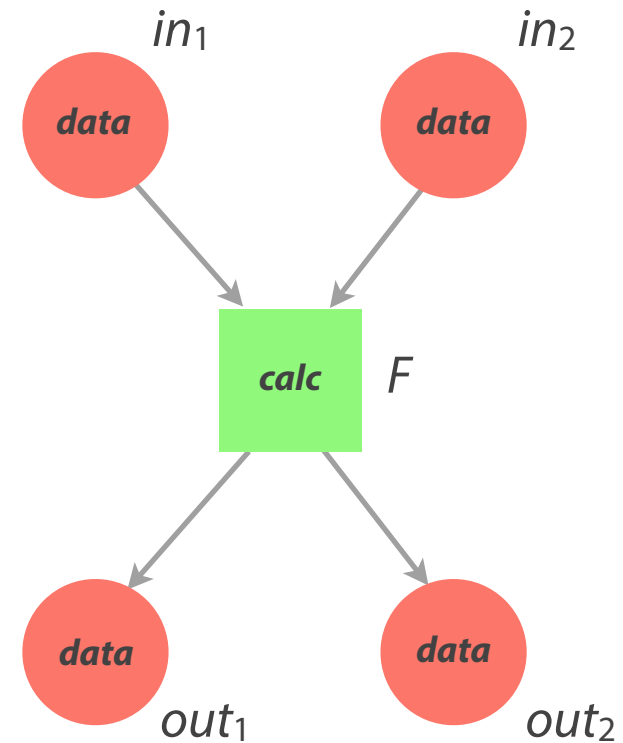
- Inspiration from the *open provenance model*

- **Any calculation: a function**, converting inputs to outputs:

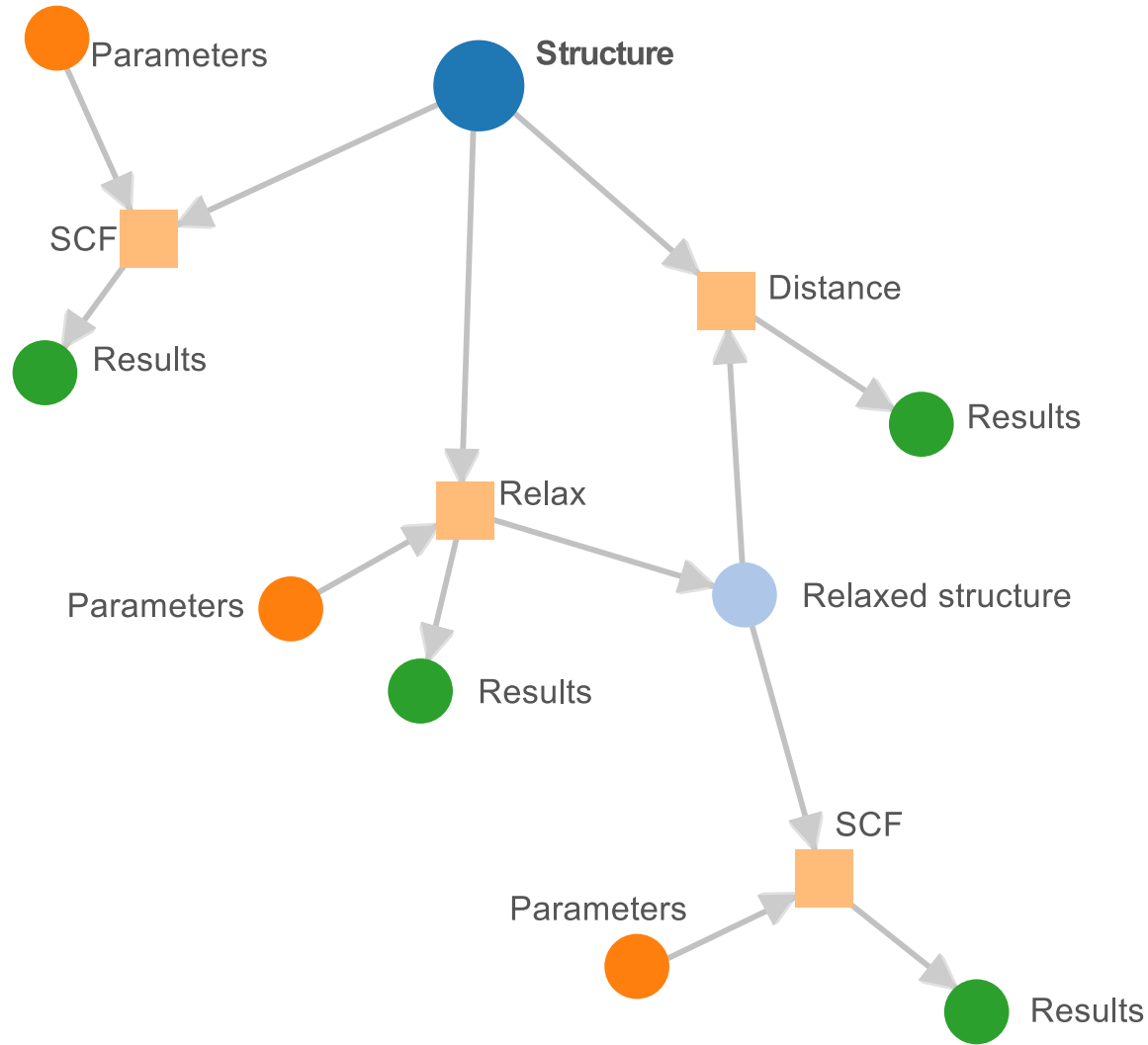
$$out_1, out_2 = F(in_1, in_2)$$

- **Each object is a node in a graph**, connected by directional labeled links

- Output nodes can be used as inputs

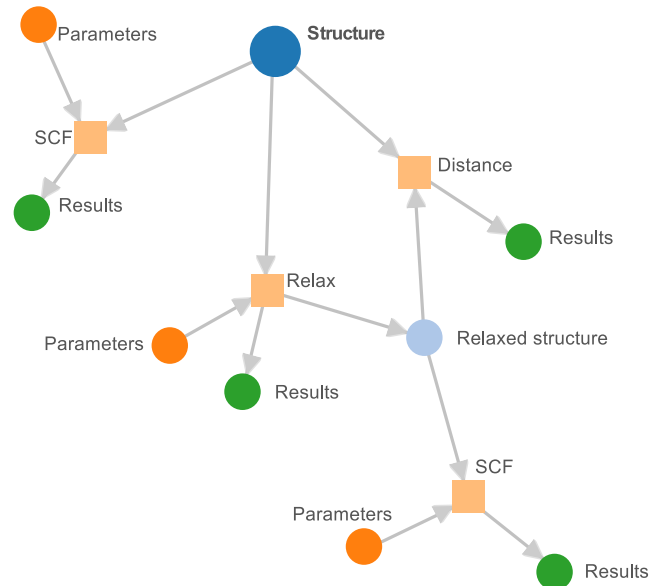


Data provenance: Directed Acyclic Graphs



Data provenance: Directed Acyclic Graphs

Directed Acyclic Graphs:
appropriate representation
of calculations, data
and their **provenance**



Data provenance: Directed Acyclic Graphs

Directed Acyclic Graphs:
appropriate representation
of calculations, data
and their **provenance**

Next questions

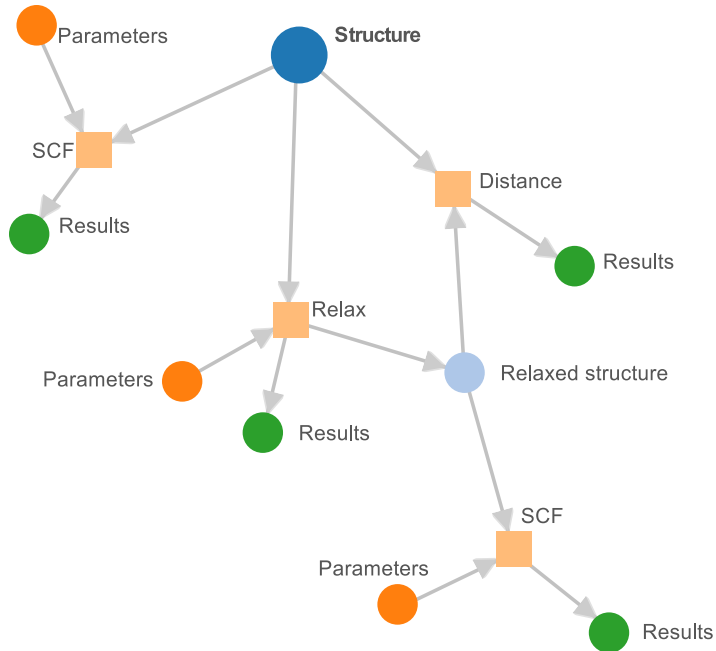
What must we store?

How can we store it?

How can we query it?



Saving the DAGs: Nodes and Links

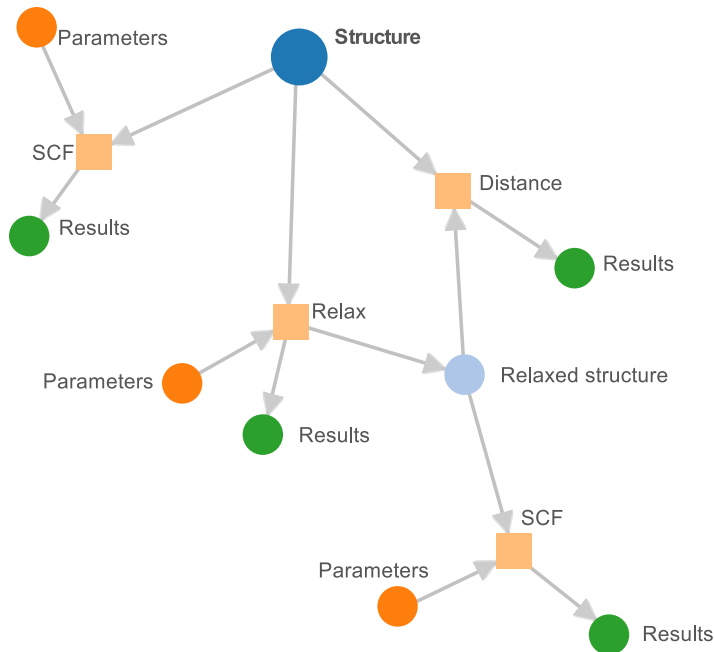


How to represent it... in SQL?

- Each **node**: row in a SQL table
 - Additional data:
 - key-value **attributes**
 - **Files/folders**
- **Links** also stored in a SQL table
⇒ *jobs provenance*



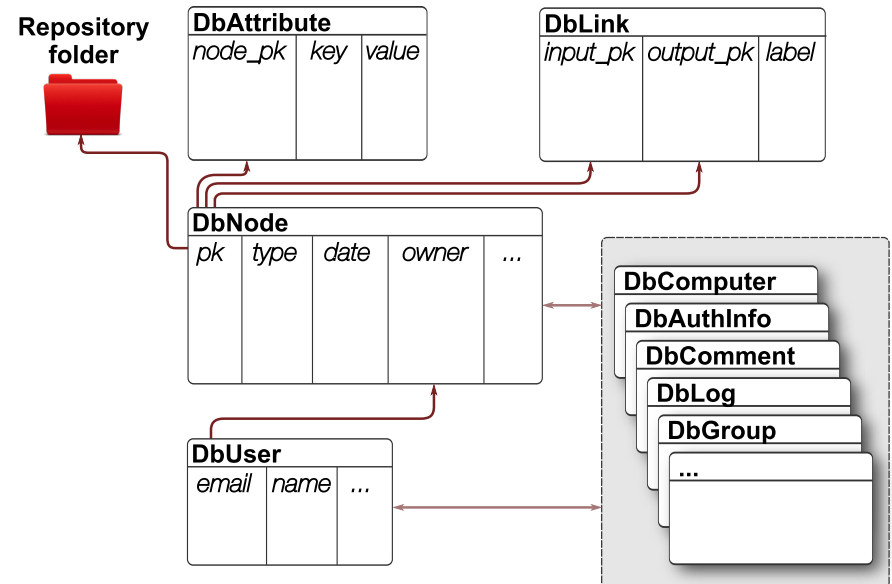
Saving the DAGs: Nodes and Links



How to represent it... in SQL?

- Each **node**: row in a SQL table
 - Additional data:
 - key-value **attributes**
 - **Files/folders**
- **Links** also stored in a SQL table
⇒ *jobs provenance*

- Multiple backends supported:
 - SQL (MySQL, PostgreSQL, ...) and *attributes EAV table via Django*
 - SQL+JSONB (PostgreSQL > 9.4) via *SQLAlchemy*
 - Easy to extend to other backends (preliminary benchmarks with *graph DBs* like *Neo4j* and *TitanDB*)

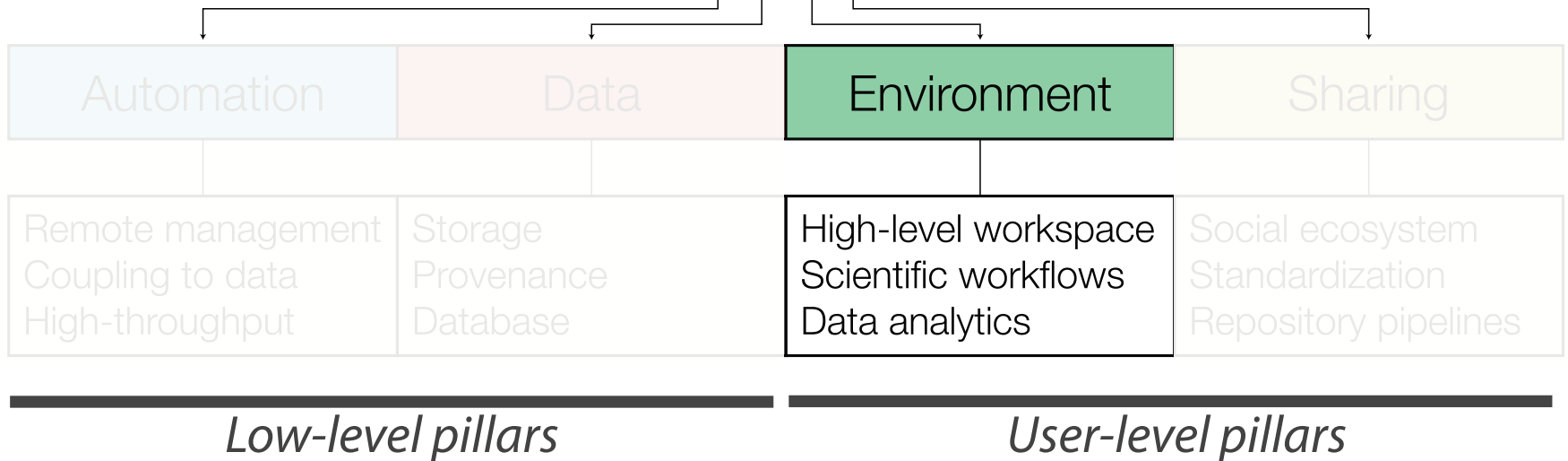


G. Pizzi et al., Comp. Mat. Sci 111,
218-230 (2016)

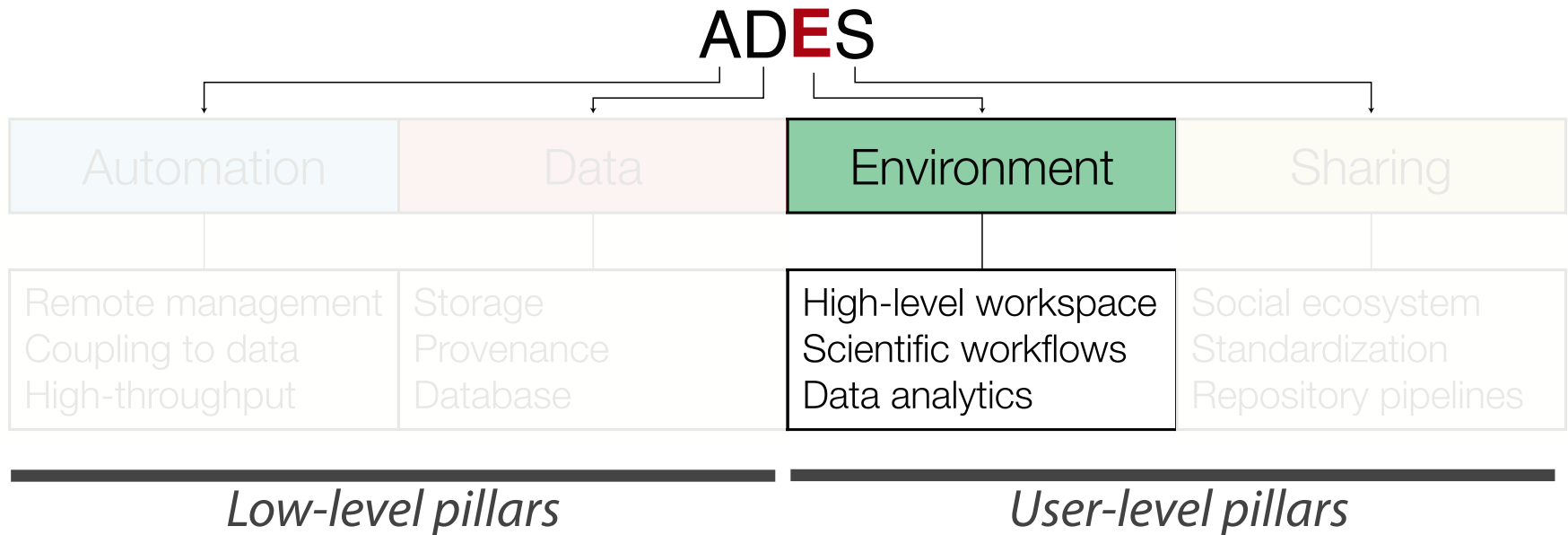


Environment in AiiDA

ADES



Environment in AiiDA

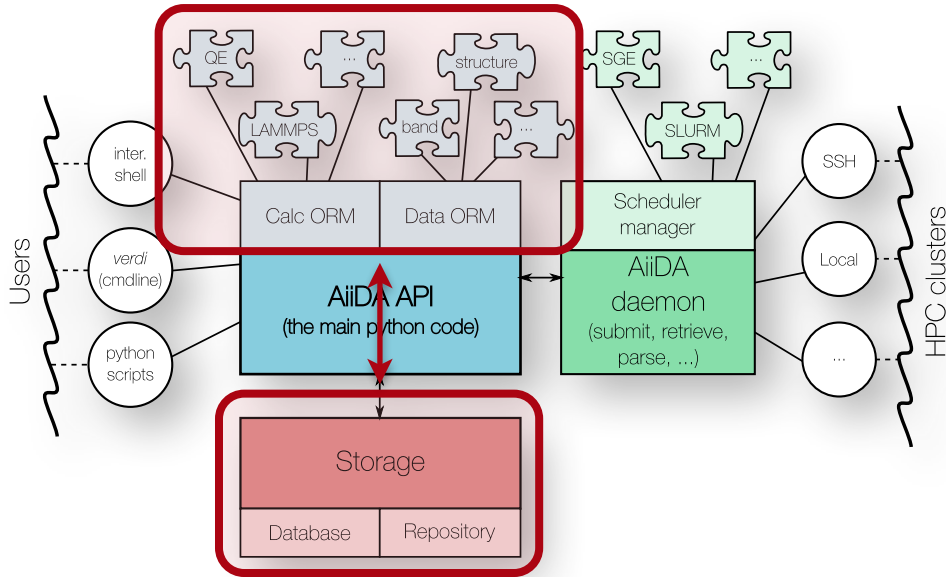


Is AiiDA easy to use?
Does it provide an advantage
over files and custom-made scripts?

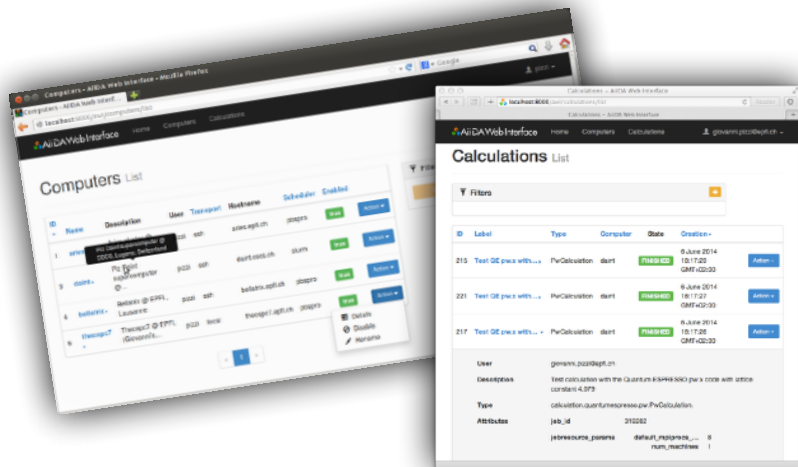


Environment in AiiDA: user interaction

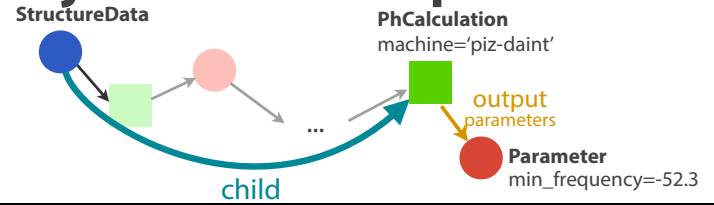
High level Python interface



Seamless user interaction



DB-agnostic query interface for AiiDA objects and their provenance



```
results = QueryBuilder(**{
    'path': [
        {'cls': PhCalculation, 'tag': 'ph'},
        {'cls': ParameterData, 'tag': 'param'}],
    'filters': {
        'ph': {'machine': "piz-daint"},
        'param': ['attributes.min_freq': {'<': -50.0}]},
    'project': {
        'param': ['attributes.min_freq'],
    })
).iterall()
```

Extensive documentation + tutorials



Reusability and modularity: AiiDA plugins



Calculation

Generation of input files for a given code

Quantum ESPRESSO, Phonopy, ASE, GPAW, Yambo, NWChem, ...



Data

Management of data objects for input/output

files&folders, parameter sets, remote data, structures, pseudos, ...



Parser

Parsing of code output & generation of new DB nodes

Quantum ESPRESSO, Phonopy, ASE, GPAW, Yambo, NWChem, ...



Transport

How to connect to a cluster

local connection, ssh, ...



Scheduler

How to interact with the scheduler

PBSPRO, Torque, SGE, SLURM, LSF, ...



Importers & exporters

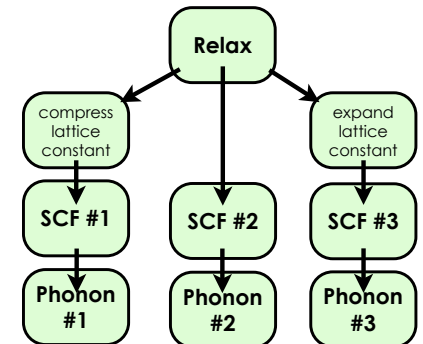
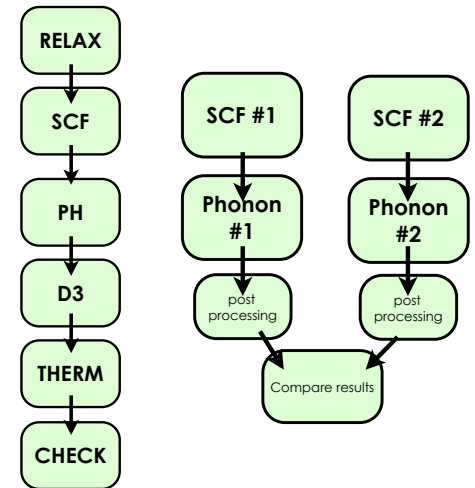
Import structures, ... from external DBs

ICSD, COD, TCOD, MPOD, ...



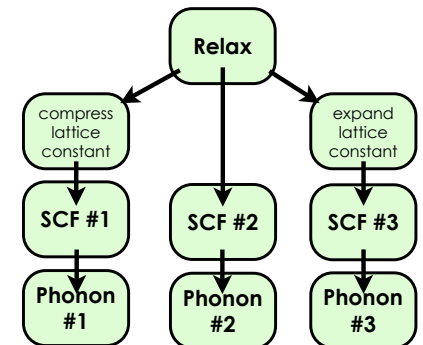
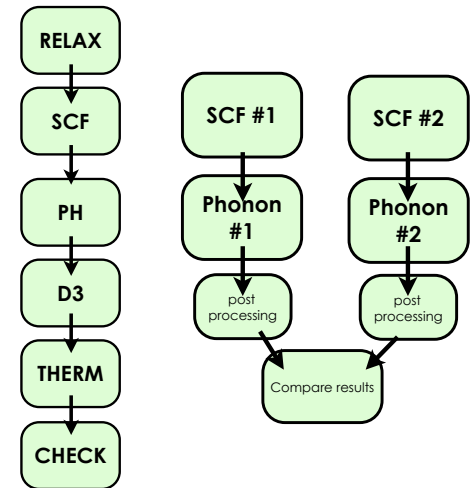
Environment in AiiDA: Scientific workflows

- Computed properties: result of **complex sequences of calculations**



Environment in AiiDA: Scientific workflows

- Computed properties: result of complex **sequences of calculations**
- ***Not only storage***: also **the management and encoding** is important for reproducibility

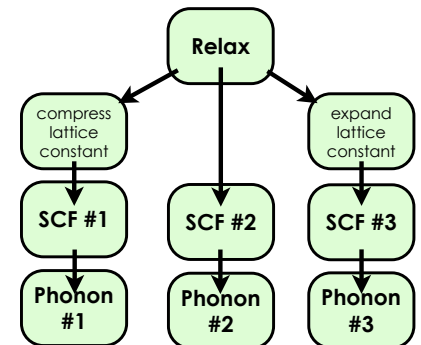
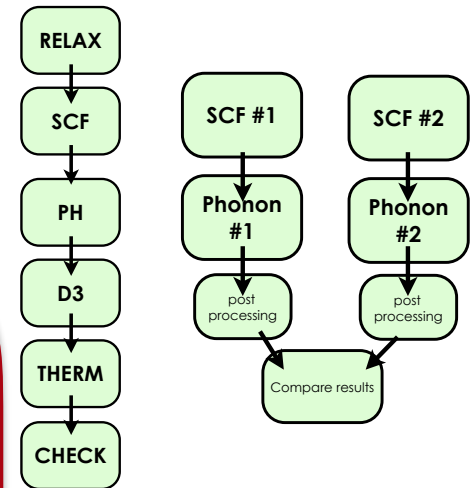


Environment in AiiDA: Scientific workflows

- Computed properties: result of complex **sequences of calculations**

- **Not only storage**: also **the management and encoding** is important for reproducibility

- We need to **encode** and to **share** “**turn-key solutions**” for:
 - the **calculation** of materials properties
 - the **automatic validation** of results



Workflows - a 'Hello World' example

```
@wf
def sum(a, b):
    return {'sum': a+b}
```

```
@wf
def prod(a, b):
    return {'prod': a*b}
```

```
@wf
def add_mul_wf(a, b, c):
    return {'result':
            prod(sum(a, b)['sum'], c)['prod']}
```

```
final_value = add_mul_wf(
    a=Int(3),
    b=Int(4),
    c=Int(5))['result']
```

$\text{final_value} = (3+4) * 5 = 35$



Workflows - a 'Hello World' example

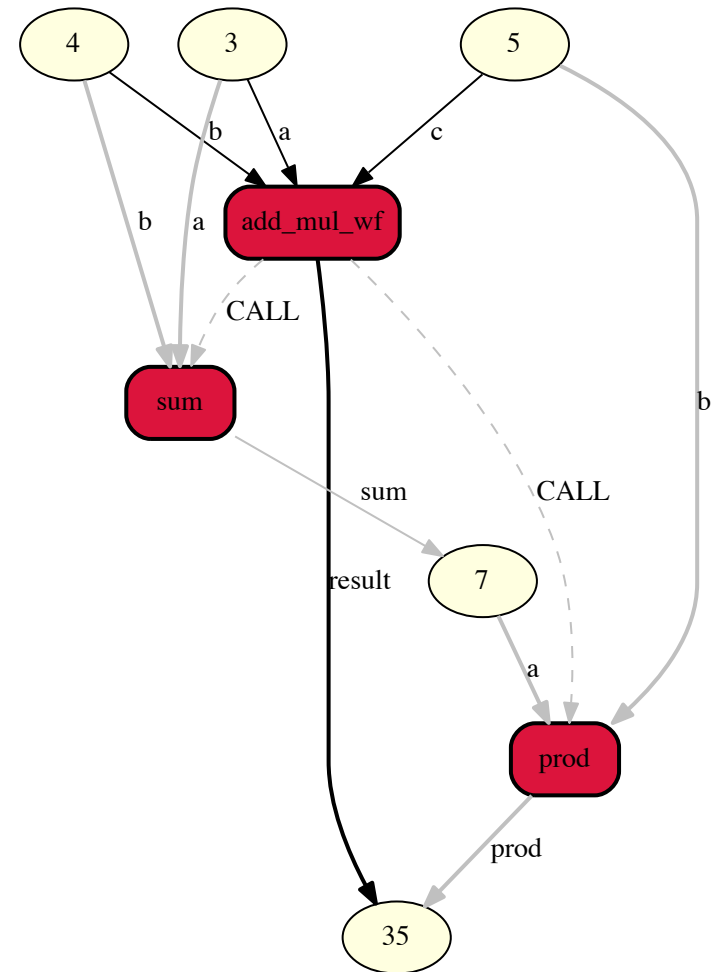
```
@wf
def sum(a, b):
    return {'sum': a+b}

@wf
def prod(a, b):
    return {'prod': a*b}

@wf
def add_mul_wf(a, b, c):
    return {'result':
        prod(sum(a, b)['sum'], c)['prod']}

final_value = add_mul_wf(
    a=Int(3),
    b=Int(4),
    c=Int(5))['result']
```

$\text{final_value} = (3+4) * 5 = 35$



Workflows - a 'Hello World' example

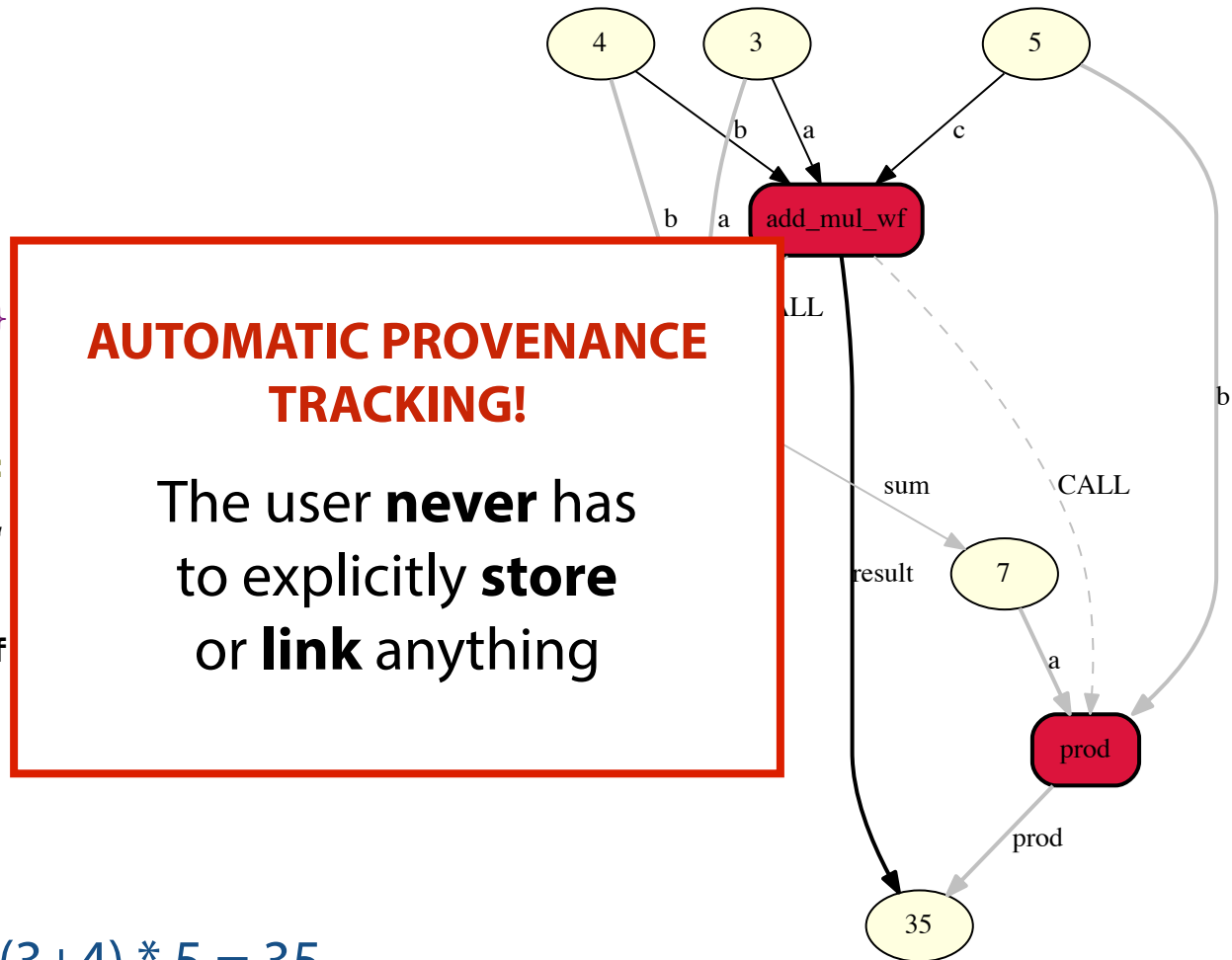
```
@wf
def sum(a, b):
    return {'sum': a+b}
```

```
@wf
def prod(a, b):
    return {'prod': a*b}
```

```
@wf
def add_mul_wf(a, b, c):
    return {'result':
        prod(sum(a, b)['
```

```
final_value = add_mul_wf
a=Int(3),
b=Int(4),
c=Int(5))['result']
```

$final_value = (3+4) * 5 = 35$



AiiDA's workflow engine features

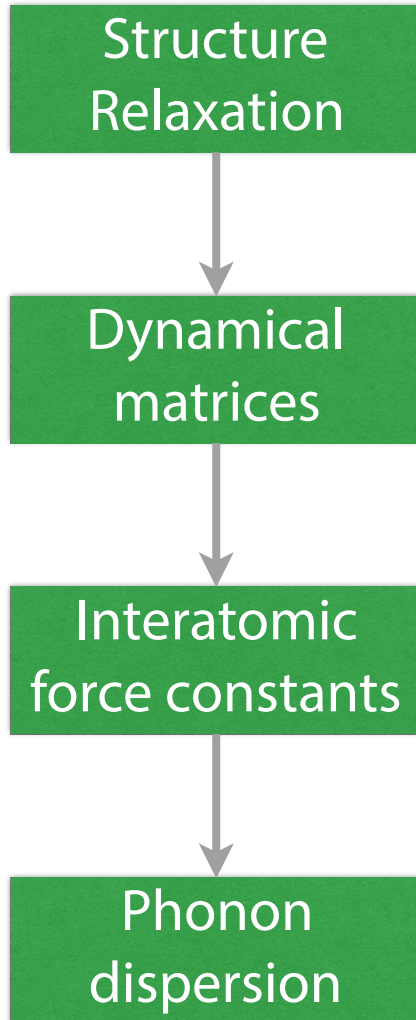
- **Workfunctions** are not a cure-all and have limitations
 - Interpreter is blocked during execution
 - Misses “sequential” character of execution
- In AiiDA we also introduce **WorkChains**:
 - Both serial and parallel execution of code and simulations
 - Checkpointing of workflows (stop/restart features)
 - Self-documenting description of inputs and outputs
 - Nesting and reuse of existing workflows





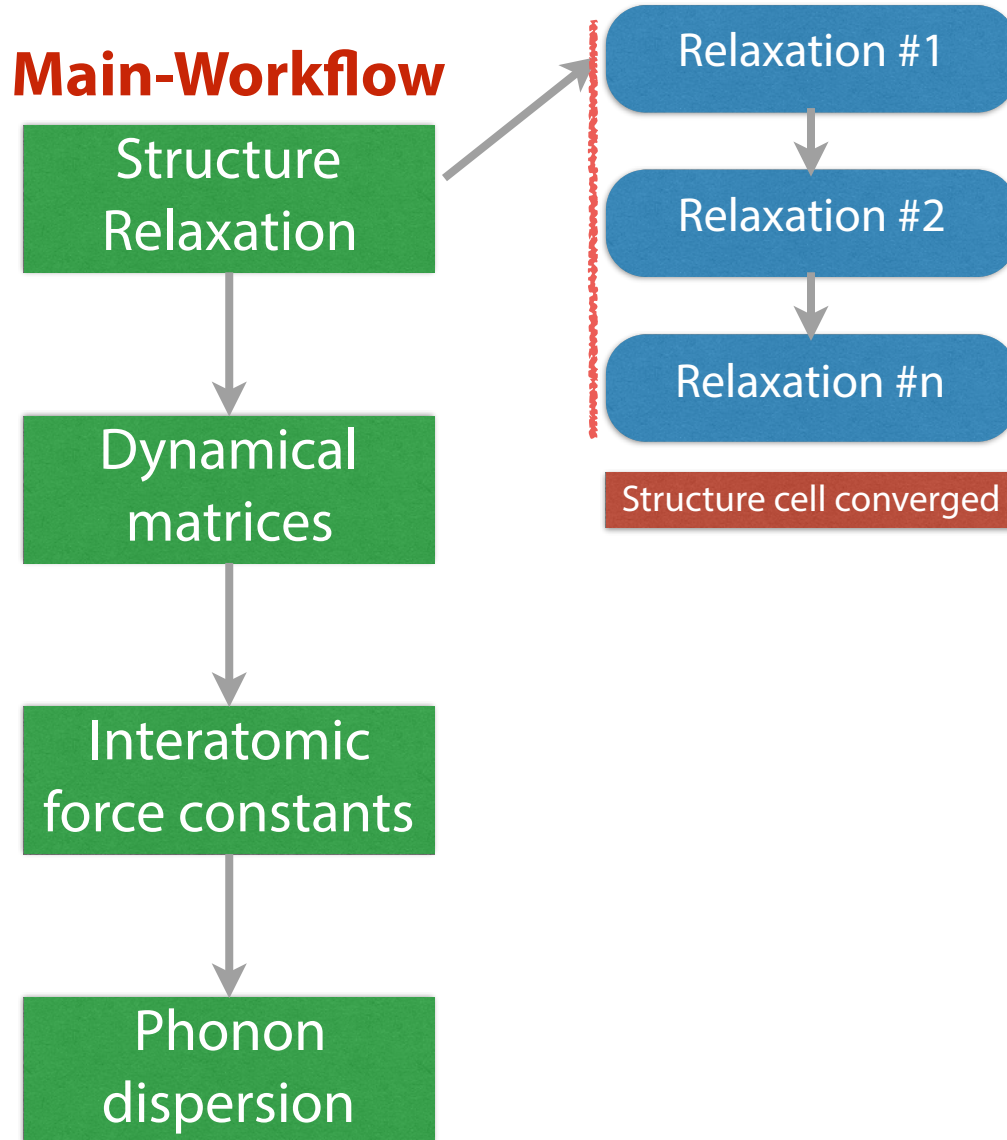
A real-life workflow example: phonon dispersions

Main-Workflow





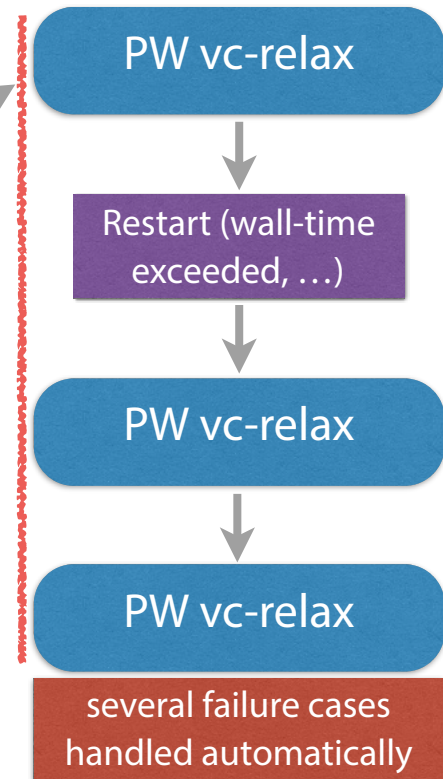
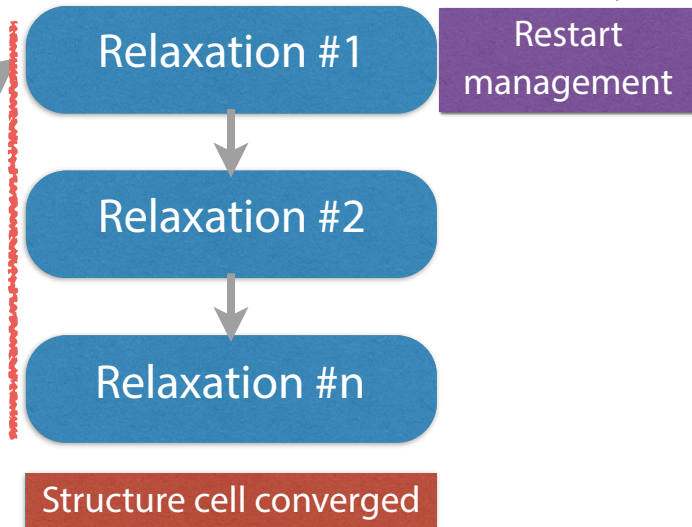
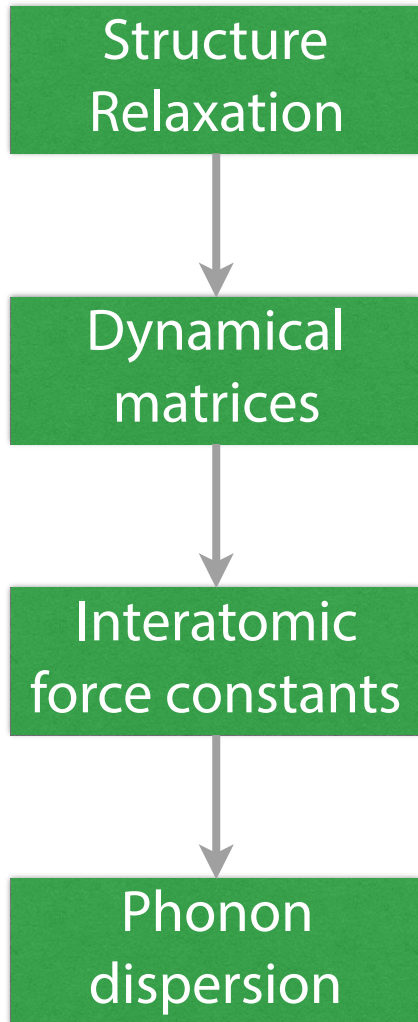
A real-life workflow example: phonon dispersions





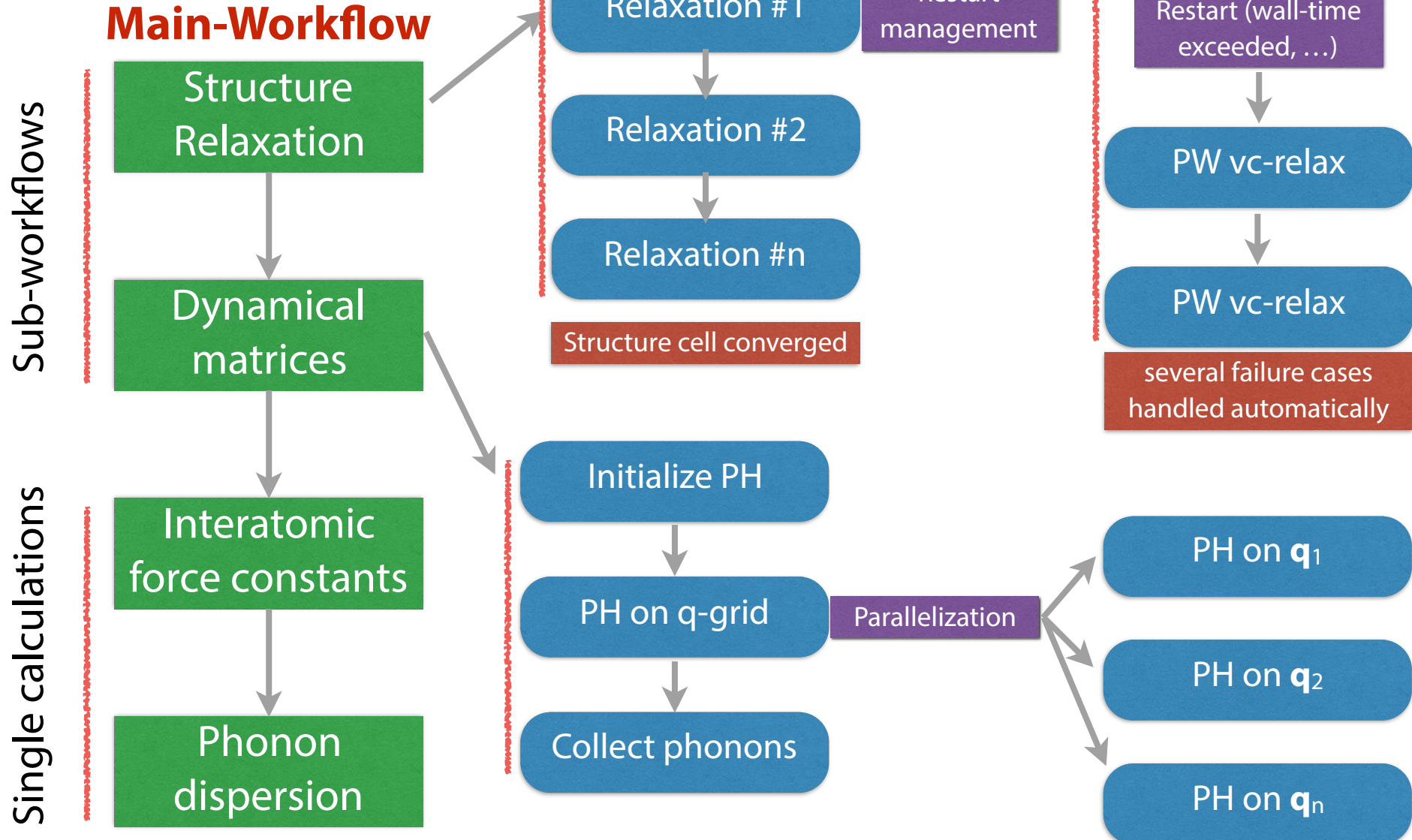
A real-life workflow example: phonon dispersions

Main-Workflow





A real-life workflow example: phonon dispersions



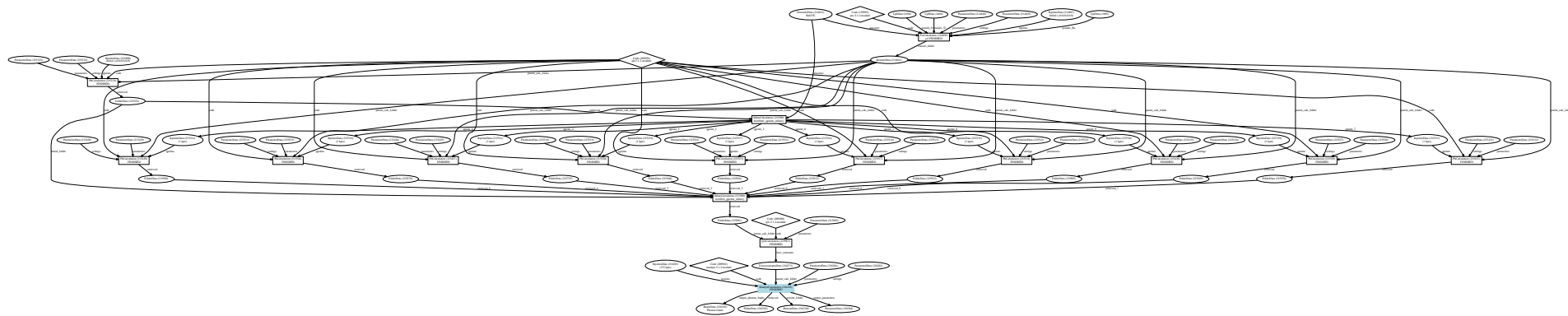
A real workflow

```
params = {'input': {'kpoints_density': 0.2,  
                  'convergence': 'tight'},  
         'structure': structure,  
         'pseudo_family': pseudo_family,  
         'machinename': 'mycluster',  
         'pw_input': {'volume_conv_threshold': 5e-2},  
         'pw_parameters': {'SYSTEM': {'ecutwfc': 30.},  
                           'ELECTRONS': {'conv_thr': 1.e-10}},  
         'ph_input': {'distance_kpoints_in_dispersion': 0.005,  
                     'diagonalization': 'cg'}}  
wf = asyncd(PhBandsWorkflow, **params)
```



A real workflow

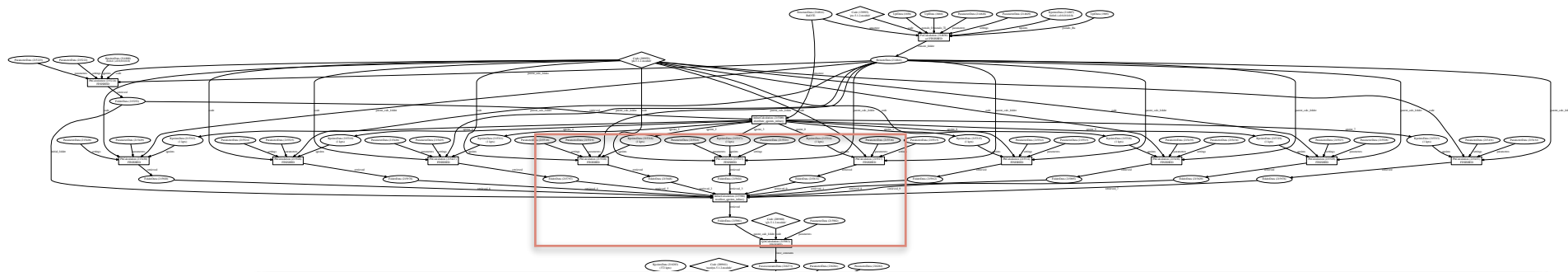
```
params = {'input': {'kpoints_density': 0.2,  
                  'convergence': 'tight'},  
         'structure': structure,  
         'pseudo_family': pseudo_family,  
         'machinename': 'mycluster',  
         'pw_input': {'volume_conv_threshold': 5e-2},  
         'pw_parameters': {'SYSTEM': {'ecutwfc': 30.},  
                           'ELECTRONS': {'conv_thr': 1.e-10}},  
         'ph_input': {'distance_kpoints_in_dispersion': 0.005,  
                      'diagonalization': 'cg'}}  
wf = asyncd(PhBandsWorkflow, **params)
```



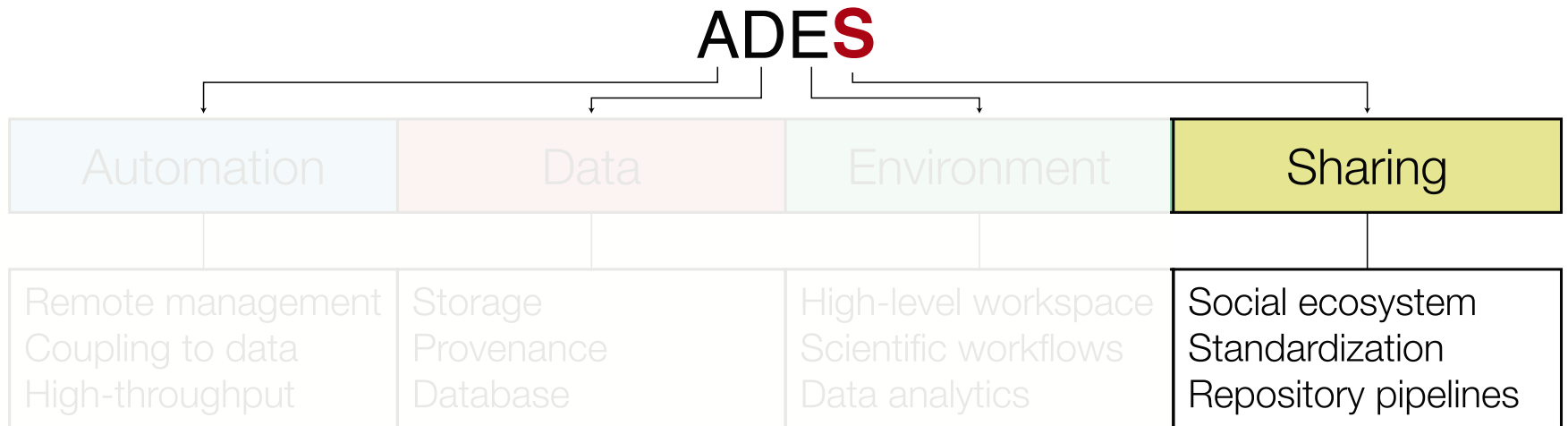
A real workflow

```
params = {'input': {'kpoints_density': 0.2,
                  'convergence': 'tight'},
         'structure': structure,
         'pseudo_family': pseudo_family,
         'machinename': 'mycluster',
         'pw_input': {'volume_conv_threshold': 5e-2},
         'pw_parameters': {
             'SYSTEM': {'ecutwfc': 30.},
             'ELECTRONS': {'conv_thr': 1.e-10}}
         'ph_input': {'distance_kpoints_in_dispersion': 0.005,
                    'diagonalization': 'cg'}}

wf = asyncd(PhBandsWorkflow, **params)
```



Sharing in AiiDA

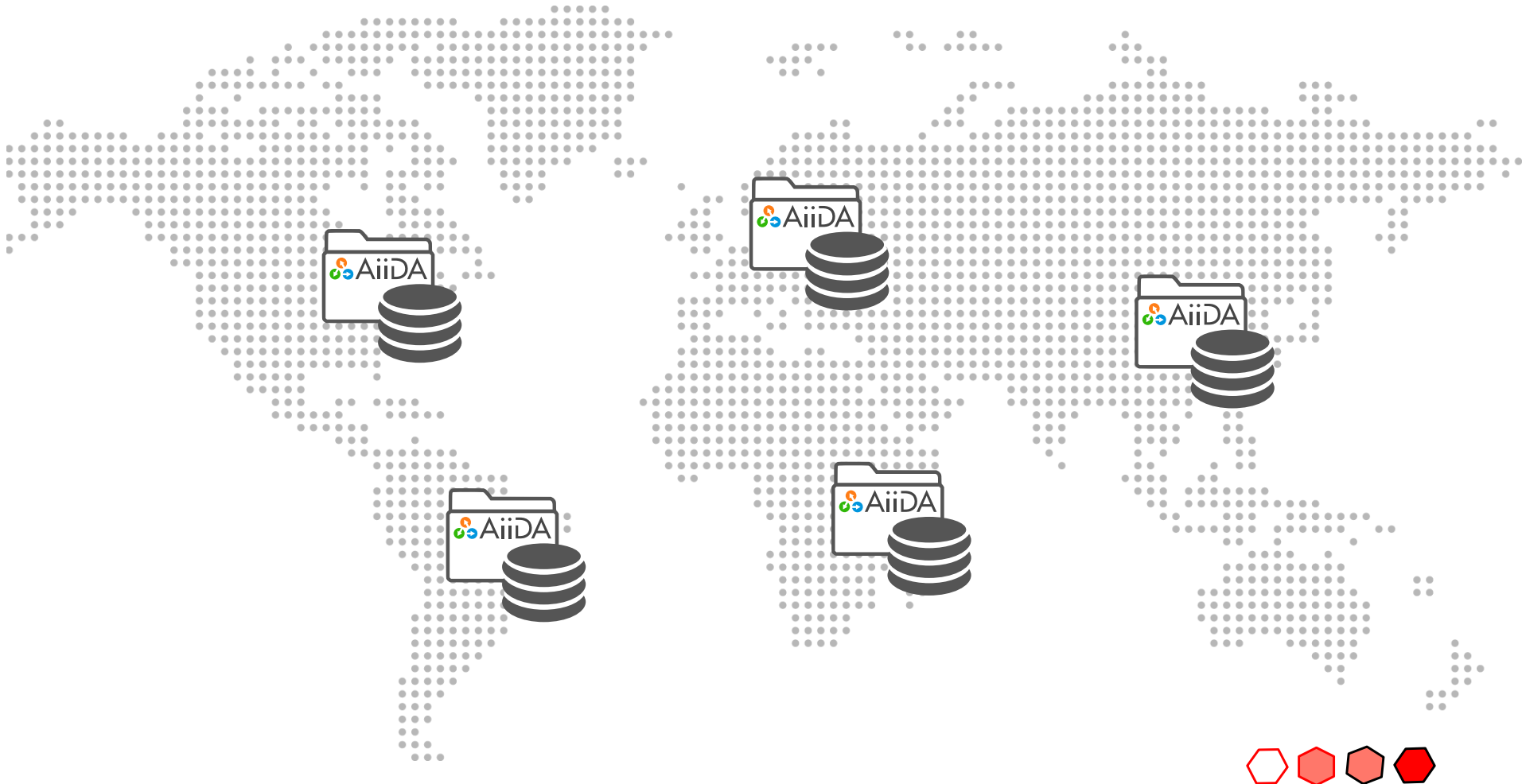


G. Pizzi et al., Comp. Mat. Sci 111, 218-230 (2016)



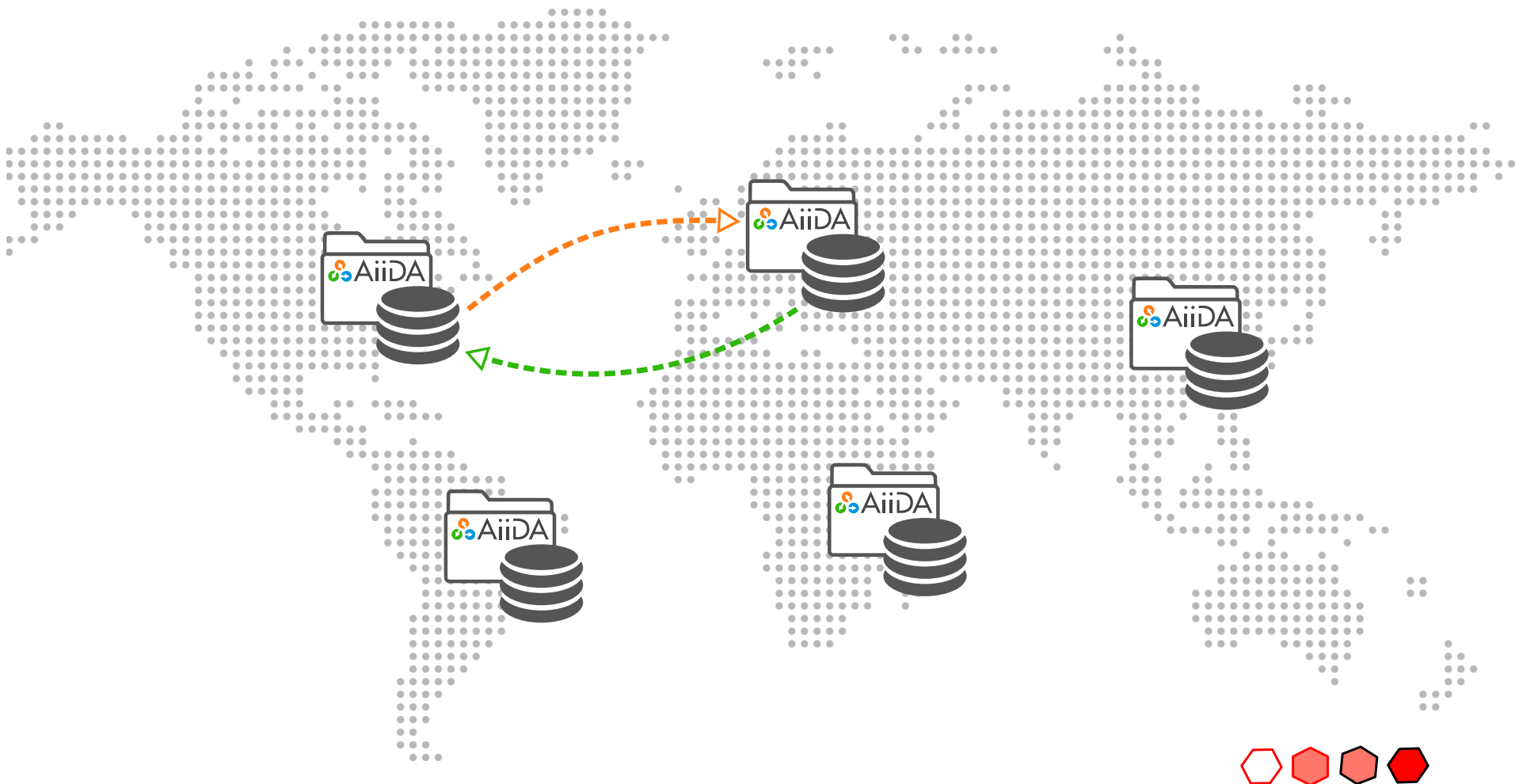
Sharing in AiiDA

- *Private* AiiDA instances
- UUIDs to uniquely identify nodes



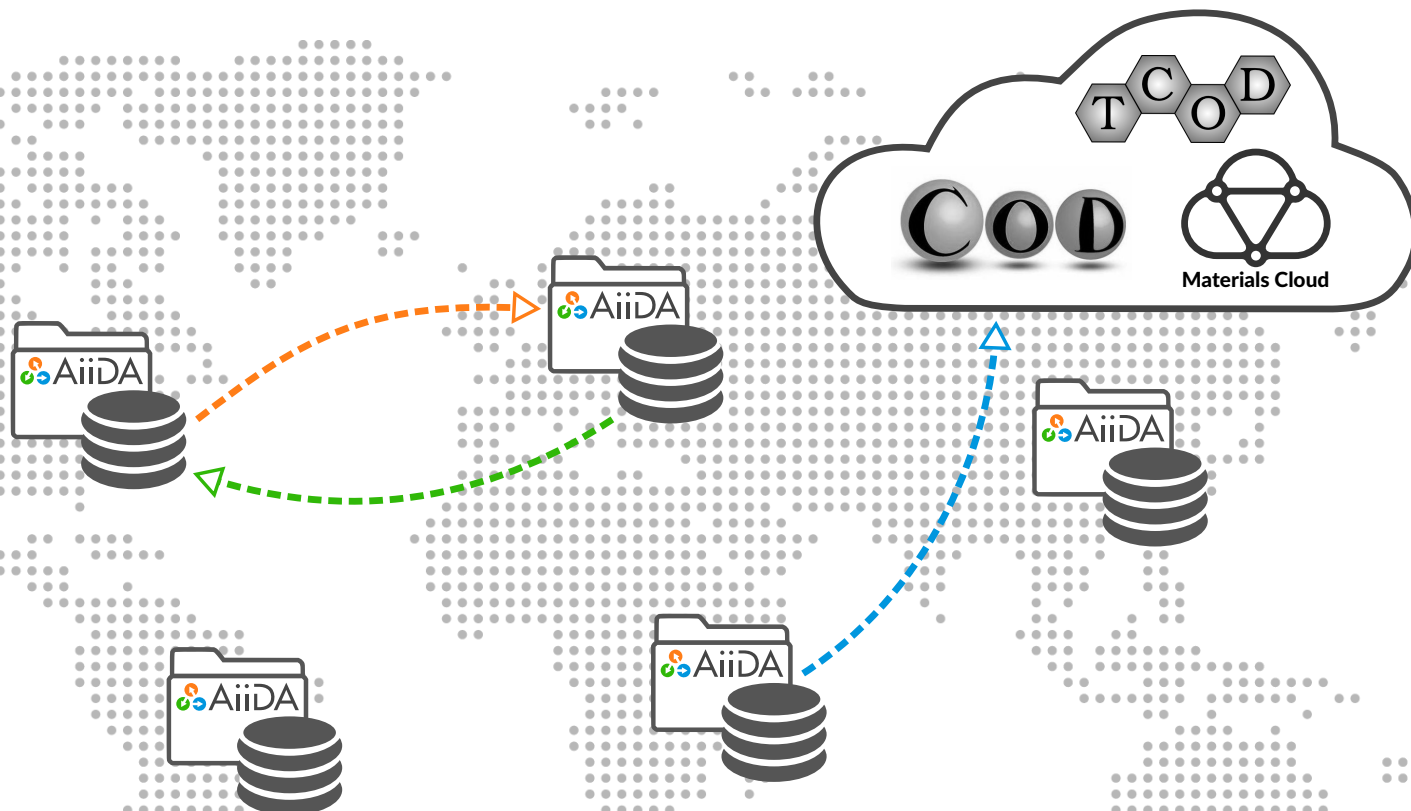
Sharing in AiiDA

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



Sharing in AiiDA

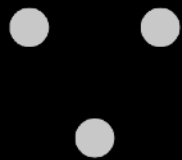
- *Private* AiiDA instances
- UUIDs to uniquely identify nodes
- Data can be shared to other AiiDA repositories **or to online repositories**



Moving to calculations on the cloud:
Materials Cloud (www.materialscloud.org)



MATERIALSCLOUD



Moving to calculations on the cloud: Materials Cloud (www.materialscloud.org)

The screenshot shows the Materials Cloud website interface. At the top, there is a navigation bar with the Materials Cloud logo and a menu with options: LEARN, WORK, DISCOVER, EXPLORE, and ARCHIVE. A blue box highlights this menu, with an arrow pointing to a larger version of the same menu overlaid on the page content. The page content is divided into several sections: a left sidebar with a navigation menu, a main content area with a breadcrumb trail (Home / Explore / -) and a node information panel, a central code editor, and a right sidebar with a diagram and a short description panel. A large white box with a black border is overlaid on the main content area, containing the text: "The sections mirror the steps of researchers through a scientific project".

MATERIALS CLOUD LEARN WORK DISCOVER EXPLORE ARCHIVE

Home / Explore / -

Node Info

Basic info

ID: 295959
UUID: 8f29333f-ed52-4394-a14d-64f429c73822
TYPE: calculation.inline.InlineCalculation.
CREATED AT: Tue, 08 Mar 2016 11:49:32 GMT

The sections mirror the steps of researchers through a scientific project

Parameter
Structure
Inline calculation
Parameter
Structure
Parameter

```
# -*- coding: utf-8 -*-  
"""  
Workflow calculate the binding energy of structures.  
"""  
  
from aiida.orm.workflow import Workflow  
from aiida.orm.calculation.inline import make_inline  
from aiida.orm import CalculationFactory, DataFactory, load_node, Group, load_workflow  
from aiida.workflows.user.epfl_theos.quantumespresso import helpers  
from aiida.workflows.user.epfl_theos.quantumespresso.pw import PWorkflow  
import time  
  
ParameterData = DataFactory('parameter')  
KpointsData = DataFactory('array.kpoints')
```

Short description

ID: 295959
TYPE: InlineCalculation
LINK TYPE
SHORT DESCRIPTION: binding_energy_inline()

VISUALISE



Moving to calculations on the cloud: Materials Cloud (www.materialscloud.org)

MATERIALS CLOUD

LEARN WORK DISCOVER EXPLORE ARCHIVE

Home / Explore / -

Node Info

Basic info

ID 295959

UUID 8f29333f-ed52-4394-a14d-64f429c73822

TYPE calculation.inline.InlineCalculation.

CREATED AT Tue, 08 Mar 2016 11:49:32 GMT

The sections mirror the steps of researchers through a scientific project

Parameter

Structure

Inline calculation

Parameter

Structure

Parameter

Structure

Parameter

Short description

ID 295959

TYPE InlineCalculation

LINK TYPE

SHORT DESCRIPTION binding_energy_inline()

VISUALISE

```
# -*- coding: utf-8 -*-
"""
Workflow calculate the binding energy of structures.
"""

from aiida.orm.workflow import Workflow
from aiida.orm.calculation.inline import make_inline
from aiida.orm import CalculationFactory, DataFactory, load_node, Group, load_workflow
from aiida.workflows.user.epfl_theos.quantumespresso import helpers
from aiida.workflows.user.epfl_theos.quantumespresso.pw import PwWorkflow
import time

ParameterData = DataFactory('parameter')
KpointsData = DataFactory('array.kpoints')
```



Moving to calculations on the cloud: Materials Cloud (www.materialscloud.org)

The screenshot shows the Materials Cloud website interface. At the top, a navigation bar contains the Materials Cloud logo and a menu with options: LEARN, WORK, DISCOVER, EXPLORE, and ARCHIVE. A red box highlights the ARCHIVE option. Below the navigation bar, a sidebar on the left lists various data categories such as Calculations, Data, Structure Data, Parameter Data, Upf Data, Array Data, Kpoints Data, Bands Data, Remote Data, Computers, and Codes. The main content area displays a node detail for an inline calculation, including its ID (295959), UUID, type (calculation.inline.InlineCalculation), and creation date (Tue, 08 Mar 2016 11:49:32 GMT). A large text box in the center of the page reads: "The sections mirror the steps of researchers through a scientific project". To the right of this text is a diagram showing a central red circle labeled "Inline calculation" connected to several blue circles labeled "Parameter" and "Structure". Below the diagram is a "Short description" section with fields for ID (295959), TYPE (InlineCalculation), LINK TYPE, and SHORT DESCRIPTION (binding_energy_inline()). At the bottom right, there is a "VISUALISE" button. The bottom of the page features a decorative row of four red hexagons.

MATERIALS CLOUD

LEARN WORK DISCOVER EXPLORE ARCHIVE

Home / Explore / -

Node Info

Basic info

LEARN WORK DISCOVER EXPLORE ARCHIVE

ID 295959

UUID 8f29333f-ed52-4394-a14d-64f429c73822

TYPE calculation.inline.InlineCalculation.

CREATED AT Tue, 08 Mar 2016 11:49:32 GMT

The sections mirror the steps of researchers through a scientific project

Parameter

Structure

Parameter

Structure

Parameter

Parameter

Structure

Structure

Parameter

Parameter

Short description

ID 295959

TYPE InlineCalculation

LINK TYPE

SHORT DESCRIPTION binding_energy_inline()

VISUALISE

```
# -*- coding: utf-8 -*-  
"""  
Workflow calculate the binding energy of structures.  
"""  
  
from aiida.orm.workflow import Workflow  
from aiida.orm.calculation.inline import make_inline  
from aiida.orm import CalculationFactory, DataFactory, load_node, Group, load_workflow  
from aiida.workflows.user.epfl_theos.quantumespresso import helpers  
from aiida.workflows.user.epfl_theos.quantumespresso.pw import PWorkflow  
import time  
  
ParameterData = DataFactory('parameter')  
KpointsData = DataFactory('array.kpoints')
```



Moving to calculations on the cloud: Materials Cloud (www.materialscloud.org)

MATERIALS CLOUD LEARN WORK DISCOVER **EXPLORE** ARCHIVE

Home / Explore / -

Calculations

- Quantum ESPRESSO - PW
- Quantum ESPRESSO - PH
- Wannier90
- Inline

Data

- Structure Data
- Parameter Data
- Upf Data
- Array Data
 - Kpoints Data
 - Bands Data
 - Remote Data

Computers

- EPFL
- CSCS
- Others
- Codes

Node Information

Basic information

ID: 295959
UUID: 8f29333f-ed52-4394-a14d-64f429c73822
TYPE: calculation.inline.InlineCalculation.
CREATED AT: Tue, 08 Mar 2016 11:49:32 GMT
MODIFIED AT: Thu, 13 Apr 2017 12:00:38 GMT
USER EMAIL: aiida@theosrv2.epfl.ch

Node Metadata

FIRST_LINE_SOURCE_CODE: 25
FUNCTION_NAME: binding_energy_inline

Source file

```
# -*- coding: utf-8 -*-  
"""  
Workflow calculate the binding energy of structures.  
"""  
  
from aiida.orm.workflow import Workflow  
from aiida.orm.calculation.inline import make_inline  
from aiida.orm import CalculationFactory, DataFactory, load_node, Group, load_workflow  
from aiida.workflows.user.epfl_theos.quantumespresso import helpers  
from aiida.workflows.user.epfl_theos.quantumespresso.pw import PWorkflow  
import time  
  
ParameterData = DataFactory('parameter')  
KpointsData = DataFactory('array.kpoints')
```

Provenance

```
graph TD  
  P1((Parameter)) --> IC((Inline calculation))  
  P2((Parameter)) --> IC  
  P3((Parameter)) --> IC  
  S1((Structure)) --> IC  
  P4((Parameter)) --> IC  
  P5((Parameter)) --> IC  
  S2((Structure)) --> IC  
  P6((Parameter)) --> IC
```

Short description

ID: 295959
TYPE: InlineCalculation
LINK TYPE:
SHORT DESCRIPTION: binding_energy_inline()

VISUALISE



Moving to calculations on the cloud: Materials Cloud (www.materialscloud.org)

The screenshot displays the Materials Cloud web interface. At the top, the navigation bar includes 'LEARN', 'WORK', 'DISCOVER', 'EXPLORE' (highlighted), and 'ARCHIVE'. The main content area is divided into several panels:

- Calculations:** A list of calculation types including Quantum ESPRESSO - PW, Quantum ESPRESSO - PH, Wannier90, Inline, Data, Structure, Parameter, Upf Data, Array, Kp, B, Rem, Cor, EP, C, C, C.
- Node Information:** A detailed view of a calculation node (ID: 295959) showing basic information such as UUID, type (Data.array.bands.banddata), creation and modification dates, and user ID.
- Band graph:** A plot showing energy bands along the high-symmetry path $\Gamma \rightarrow M; M \rightarrow K; K \rightarrow \Gamma; \Gamma \rightarrow M; M \rightarrow \Gamma$. The energy ranges from -7 to 4 eV.
- Provenance:** A diagram showing the workflow of the calculation, including inputs like Quantum Espresso pw and outputs like bands.
- Short description:** A section providing metadata for the node, including ID (295959), type (InlineCalculation), link type, and a short description (binding_energy_inline()).
- Code:** A snippet of Python code for loading and working with the node:

```
from materialscloud import load_node, Group, load_workflow
import helpers
pw import PWorkflow
```

At the bottom right, there is a 'VISUALISE' button and a decorative graphic of four colored hexagons (white, red, orange, black).

The challenge of data curation (Discover section)



MATERIALS
CLOUD

LEARN

WORK

DISCOVER

EXPLORE

ARCHIVE



Home

ThermoChemicals

2D Structures

Sketch-map

Home / Discover / Menu

Discover home

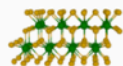
This section will contain a curated set of results including structures and their properties as generated by NCCR members.



Thermochemical properties of solids

Authors: Nicolas Hoermann, Martin Uhrin (THEOS)

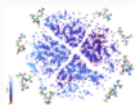
Description: High-throughput study of binary compounds to derive the temperature dependence on the free energy, enthalpy and volumetric thermal expansion



2D structures and layered materials

Authors: Nicolas Mounet, Marco Gibertini, Philippe Schwaller, Giovanni Pizzi (THEOS)

Description: Results from screening all known 3D crystal structures finding those that can be computationally exfoliated producing 2D materials candidates



Sketch-map

Authors: Michele Ceriotti, Sandip De

Description: An intuitive visualization of a set of atomic structures can be obtained using Sketch-map



The challenge of data curation (Discover section)

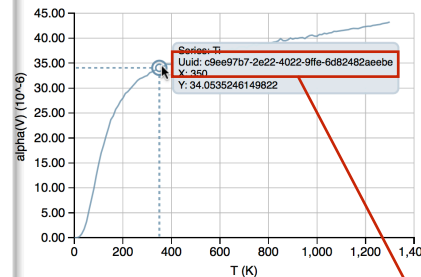
Compound: Ti

Curated datasets

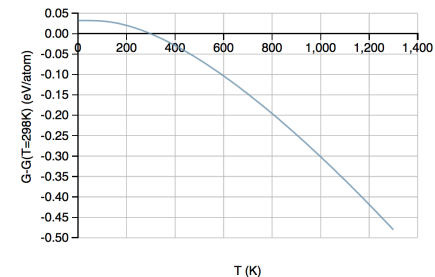
After raw data generation, but before interpretation in a paper

Interactive versions of figures on a paper, with **direct access to calculation provenance!**

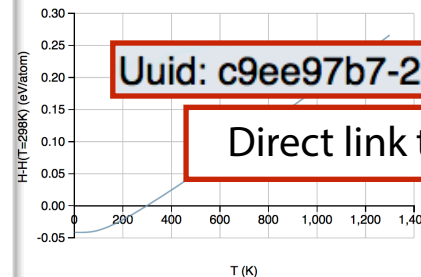
Thermal Expansion Coefficient vs Temperature (quasi harmonic Debye)



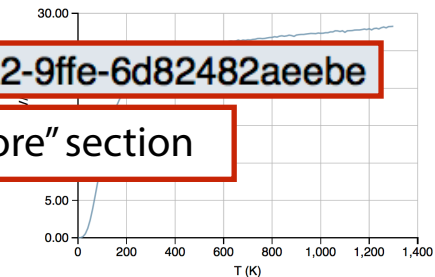
Gibbs Free Energy vs Temperature (quasi harmonic Debye)



Enthalpy vs Temperature (quasi harmonic Debye)



Heat Capacity vs Temperature (quasi harmonic Debye)



Uuid: c9ee97b7-2e22-4022-9ffe-6d82482aeebe

Direct link to "Explore" section



Work section: App-store model



MATERIALS
CLOUD

LEARN

WORK

DISCOVER

EXPLORE

ARCHIVE

“App-store”-like model for plugins
& workflows, e.g.

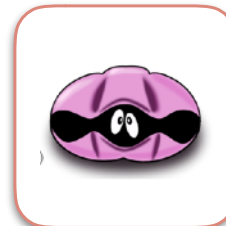
- **Computers**
- **Importers/exporters**
- **Calculations plugins**
- **Workflows & turn-key solutions**
- ...

The MaterialsCloud AiiDA app-store

Simulation codes



Quantum
ESPRESSO



Yambo



CP2K



Fleur

Computing clusters



CSCS (CH)



CINECA (IT)

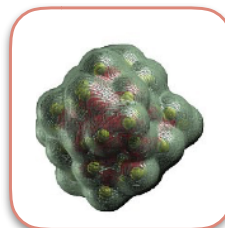


BSC (ES)

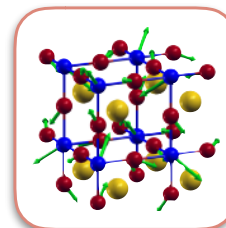


Jülich (DE)

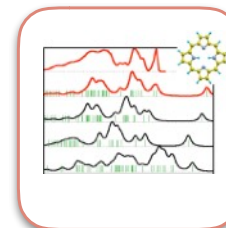
Turn-key workflows



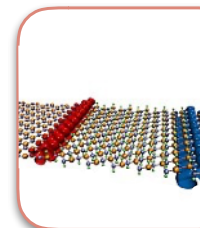
Electronic
density



Phonons



Optical
properties



Transport

Work section: App-store model



MATERIALS
CLOUD

LEARN

WORK

DISCOVER

EXPLORE

ARCHIVE

The MaterialsCloud AiiDA app-store

“App-store”-
& workflows,

- **Computer**
- **Importers**
- **Calculatio**
- **Workflow**
- ...

But also:
“**AiiDA on the cloud**”
for a **data-on-demand model**
with **spot virtual machines** in OpenStack
Running a full stack including AiiDA, database,
object store, connection to HPC

Powered by **open-source software**
and **robust AiiDA workflows**

With intuitive **GUI customisable** by users
(With interactive JupyterLab notebooks)
work in collab. with Ole Schütt, EMPA, CH



K



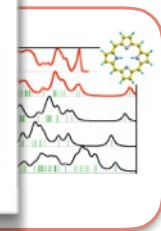
Fleur



(ES)



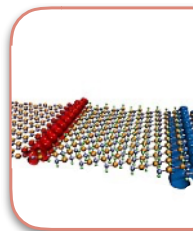
Jülich (DE)



Electronic
density

Phonons

Optical
properties



Transport

Acknowledgements

The AiiDA team



Giovanni
Pizzi
(EPFL)



Andrea
Cepellotti
(EPFL)



Nicolas
Mounet
(EPFL)



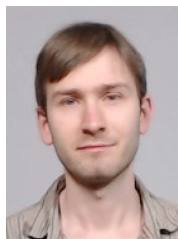
Fernando
Gargiulo
(EPFL)



Boris
Kozinsky
(BOSCH)



Nicola
Marzari
(EPFL)



Martin
Uhrin
(EPFL)



Leonid
Kahle
(EPFL)



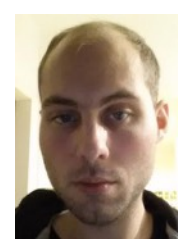
Spyros
Zoupanos
(EPFL)



Snehal
Waychal
(EPFL)



Sebastiaan
Huber
(EPFL)



Rico
Häuselmann
(EPFL)



Andrius
Merkys
(Vilnius)



Acknowledgements

The AiiDA team



Giovanni
Pizzi
(EPFL)



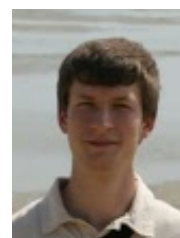
Andrea
Cepellotti
(EPFL)



Nicolas
Mounet
(EPFL)



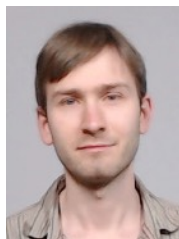
Fernando
Gargiulo
(EPFL)



Boris
Kozinsky
(BOSCH)



Nicola
Marzari
(EPFL)



Martin
Uhrin
(EPFL)



Leonid
Kahle
(EPFL)



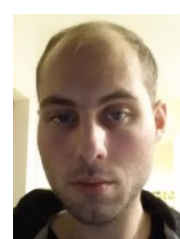
Spyros
Zoupanos
(EPFL)



Snehal
Waychal
(EPFL)



Sebastiaan
Huber
(EPFL)



Rico
Häuselmann
(EPFL)



Andrius
Merkys
(Vilnius)

Former AiiDA team members — Riccardo Sabatini, EPFL — **Plugin contributors** — **Quantum ESPRESSO NEB:** Marco Gibertini (EPFL); **Quantum ESPRESSO DOS, PDOS; Wannier90:** Daniel Marchand (EPFL); **CP2K:** Aliaksandr Yakutovich (EMPA), Uli Schauer (ETHZ), Tiziano Müller, Andreas Gloess, Patrick Seewald (UZH); **FLEUR:** Jens Broeder, Gregor Michalicek, Daniel Wortmann (Jülich); **Exciting:** Anton Kozhevnikov (CSCS); **YAMBO:** Andrea Ferretti, Mike Atambo, Daniele Varsano (CNR-NANO), Gianluca Prandini, Antimo Marrazzo (EPFL); **SIESTA:** Victor Garcia Suarez (Uni Oviedo); **VASP:** Mario Zic (Trinity College Dublin); ... **Contributors** — Prof. Christoph Koch (EPFL); Valentin Bersier, Philippe Schwaller, Ivano Castelli (THEOS EPFL); Marco Dorigo (ICAMS - Bochum); Eric Hontz (MIT & Bosch RTC)

Infrastructure Projects

MARVEL NCCR

Materials' Revolution: Computational Design
and Discovery of Novel Materials

Swiss National Centre of Competence
Started May 2014, funded 2014-2026

41 PIs from 12 Institutions
(including the Swiss supercomputing
center CSCS)

<http://marvel-nccr.ch>

MaX

Driving the exascale transition:
Materials design at the eXascale


EU H2020 e-infrastructure project
From Sep 2015 to Feb 2018

Modena, Trieste, EPFL, Barcelona, Julich
+ 5 supercomputing centers
(CINECA, CSCS, BCS, Julich, KTH)

<http://max-center.eu>

The image shows two website screenshots. The left screenshot is from the MARVEL website, featuring a navigation bar with 'Home', 'About', 'Research', and 'People'. Below the navigation is a header with the MARVEL logo and a search bar. The main content area includes a large image of a building with the caption 'In collaboration with the University of Basel'. Below this is a section titled 'Who's behind MARVEL?' featuring a profile of Michele Parrinello, a professor of computational sciences at the University of Basel and ETHZ. Another section titled 'Recents News' highlights 'Piz Daint in third position of the TOP500 list of most powerful computers'. A sidebar on the left lists various organizational roles like 'Management Team' and 'Executive Committee'. The right screenshot is from the MaX website, featuring a navigation bar with 'the centre', 'materials design', 'services', 'people', 'events', 'news', 'contact', 'user portal', 'intranet', and 'industries'. The main content area has a section titled 'MaX - Materials design at the Exascale' with the subtitle 'a European centre of excellence'. The text describes MaX as a user-focused, problem-oriented European Centre of Excellence working at the frontiers of High Performance Computing (HPC) technologies. A word cloud on the right side of the page contains terms like 'technology', 'H2020', 'performance', 'platform', 'computing', 'strategic', 'development', 'design', 'news', 'project', 'launch', 'materials', 'excellence', 'European', 'e-infrastructure', 'HPC', 'exascale', 'center', 'agenda', 'research', 'modena', 'molinari', 'ETP4HPC', and 'H2020-EINFRA-2015-1'.

Conclusions

 **AiiDA** A computational science platform adopting the **ADES** model:

 **AiiDA**

Conclusions



A computational science platform adopting the **ADES** model:

Automation

automate repetitive tasks via daemon, **abstracting** into APIs



Conclusions



A computational science platform adopting the **ADES** model:

Automation

automate repetitive tasks via daemon, **abstracting** into APIs

Data

reproducibility&provenance, directed acyclic graphs, queries

Conclusions



A computational science platform adopting the **ADES** model:

Automation

automate repetitive tasks via daemon, **abstracting** into APIs

Data

reproducibility&provenance, directed acyclic graphs, queries

Environment

flexible platform; **workflows** to encode scientists' knowledge

Conclusions



A computational science platform adopting the **ADES** model:

Automation

automate repetitive tasks via daemon, **abstracting** into APIs

Data

reproducibility&provenance, directed acyclic graphs, queries

Environment

flexible platform; **workflows** to encode scientists' knowledge

Sharing

social ecosystem to encourage interactions

Contacts and info:



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

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

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



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



@aiidateam

