Managing computational materials science in the high-throughput era

The ADES model and the AiiDA infrastructure

Giovanni Pizzi (EPFL, Switzerland)

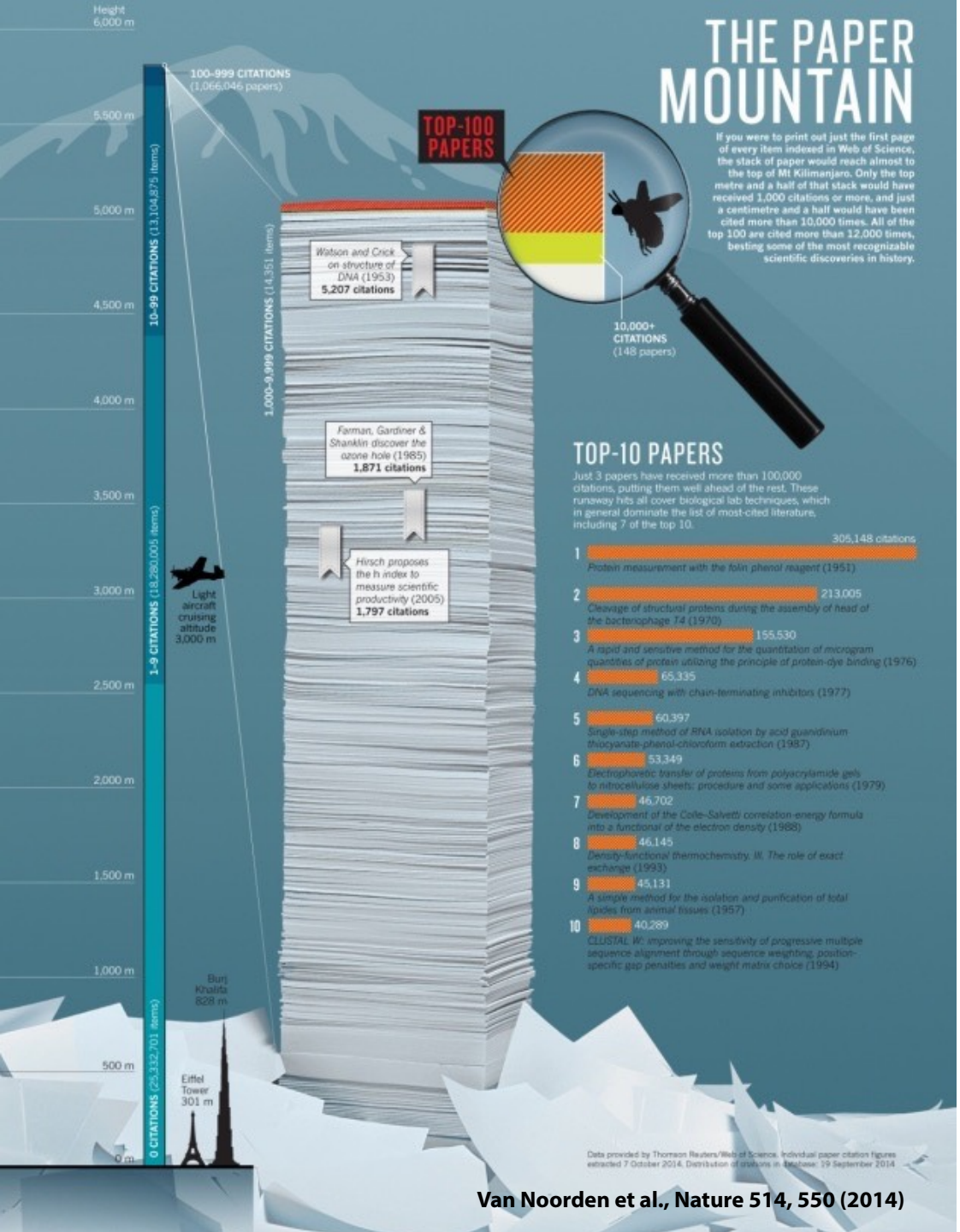
MAX DRIVING THE EXASCALE TRANSITION

MARVEL

EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Nature, November 2014

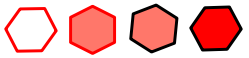
12 papers on DFT in the top-100 most cited papers in the entire scientific literature



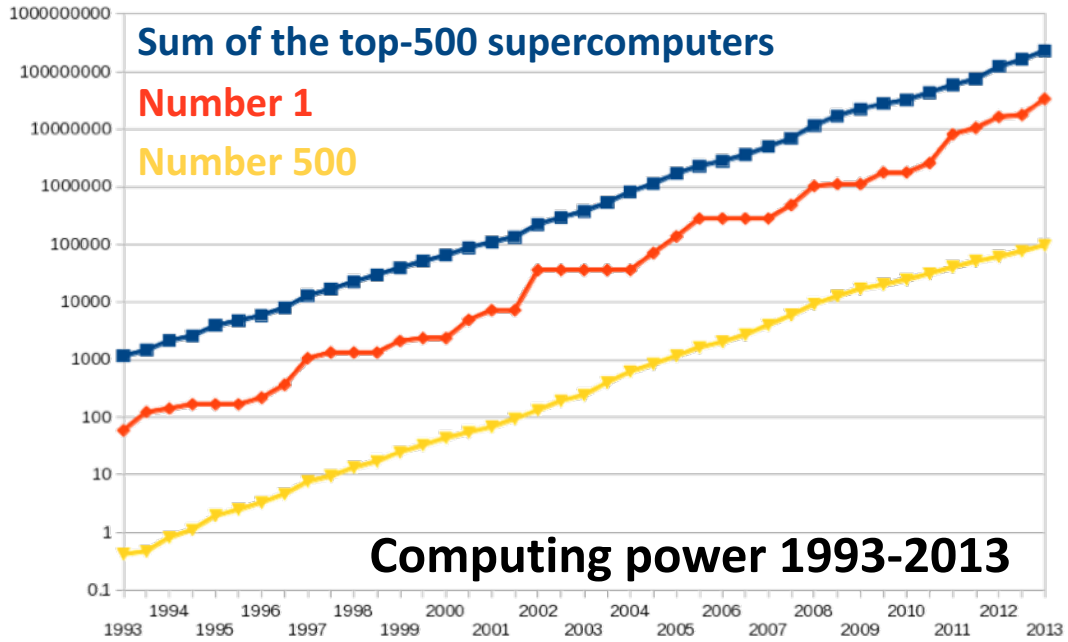
Van Noorden et al., Nature 514, 550 (2014)

Van Noorden et al., Nature 514, 550 (2014)

Accuracy and predictive power of quantum engines



Accuracy and predictive power of quantum engines



150,000x increase in the past 20 years

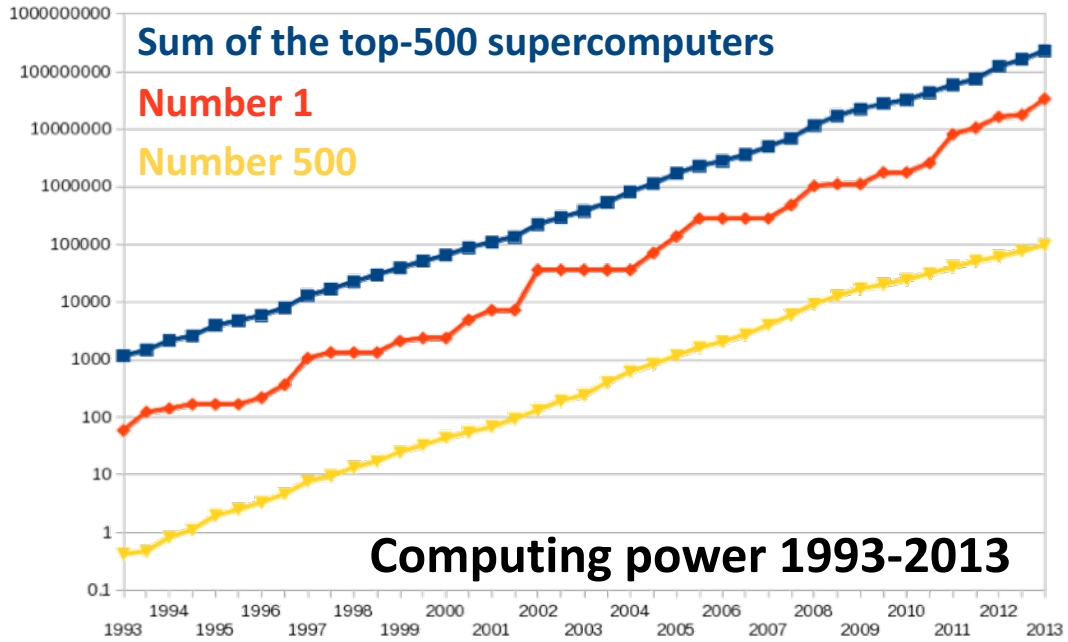
1 month (1993)

↓

10 seconds (2015)

Van Noorden et al., Nature 514, 550 (2014)

Accuracy and predictive power of quantum engines



150,000x increase in the past 20 years

1 month (1993)

↓

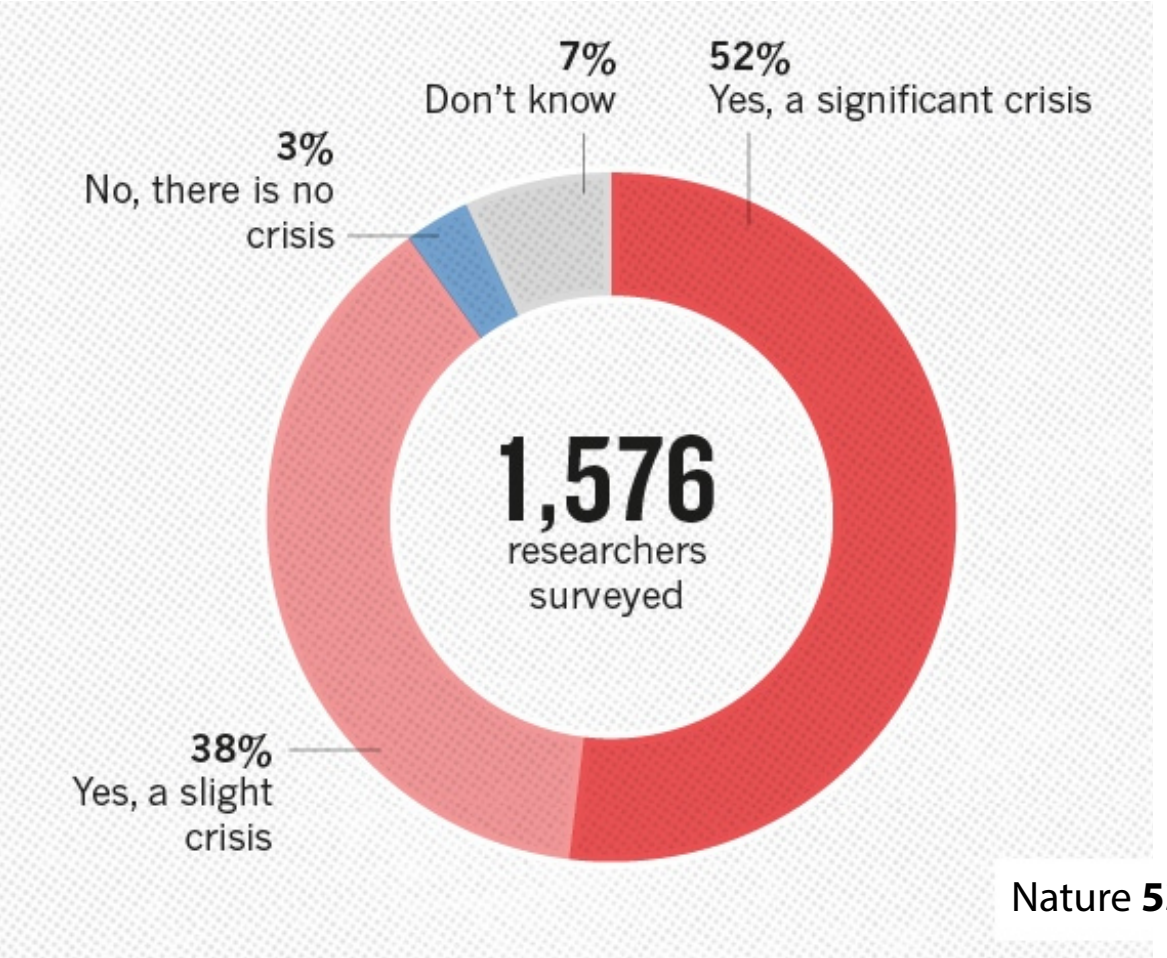
10 seconds (2015)

Result: materials design and discovery via

high-throughput computations

Reproducibility: a cornerstone of the scientific method

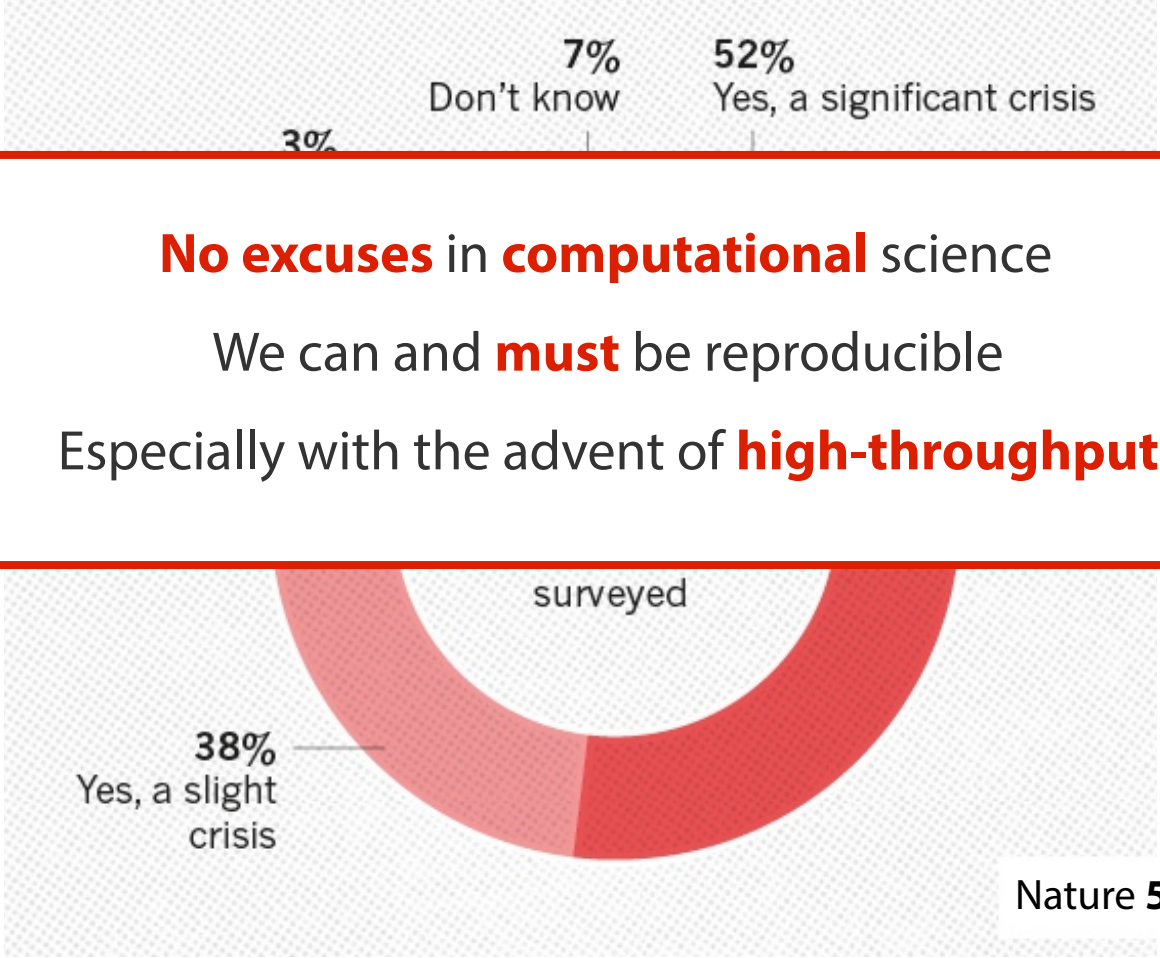
IS THERE A REPRODUCIBILITY CRISIS?



Nature **533**, 452–454 (2016)

Reproducibility: a cornerstone of the scientific method

IS THERE A REPRODUCIBILITY 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 | ˈpräv-nən(t)s, ˈprä-və-ˌnän(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 | ˈpräv-nən(t)s, ˈprä-və-ˌnän(t)s

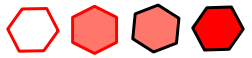
1. The origin or source of something
2. The history of ownership of a valued object or work of art or literature

To guarantee reproducibility it is not enough to just store the

WHAT

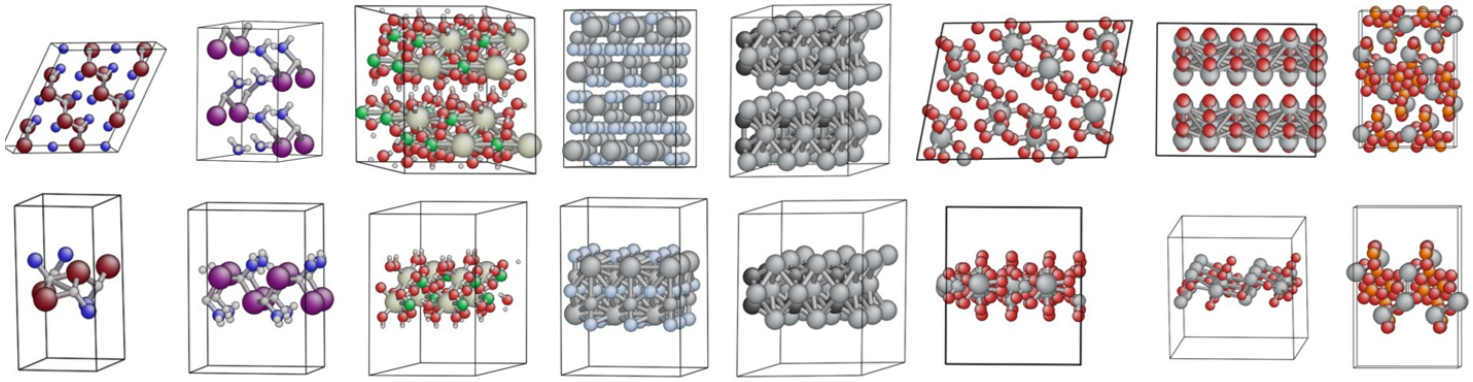
but it is crucial to also store the

HOW

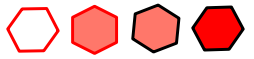


Importance of provenance: an high-throughput example

Discovering new two-dimensional materials

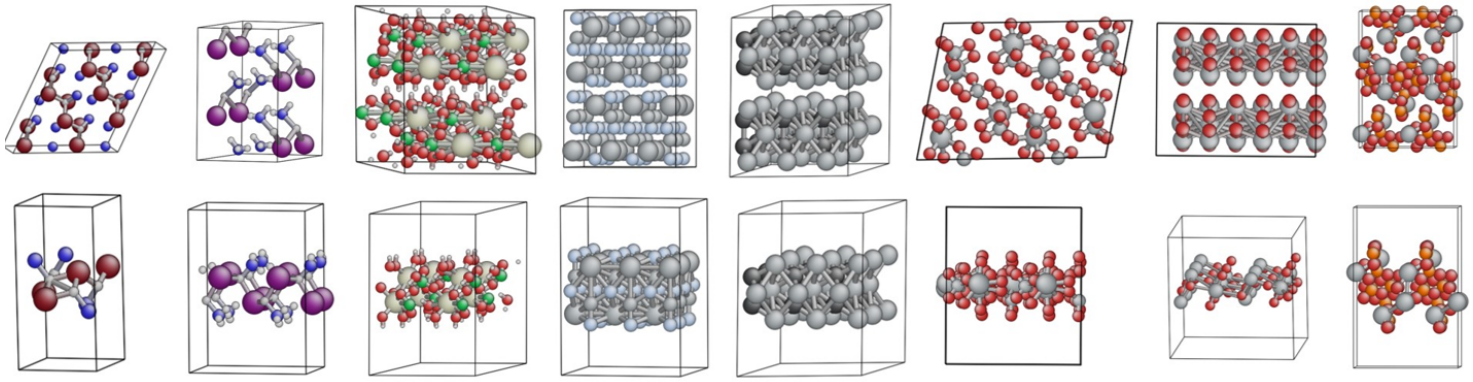


N. Mounet, GP et al., arXiv:1611.05234 (2016)



Importance of provenance: an high-throughput example

Discovering new two-dimensional materials



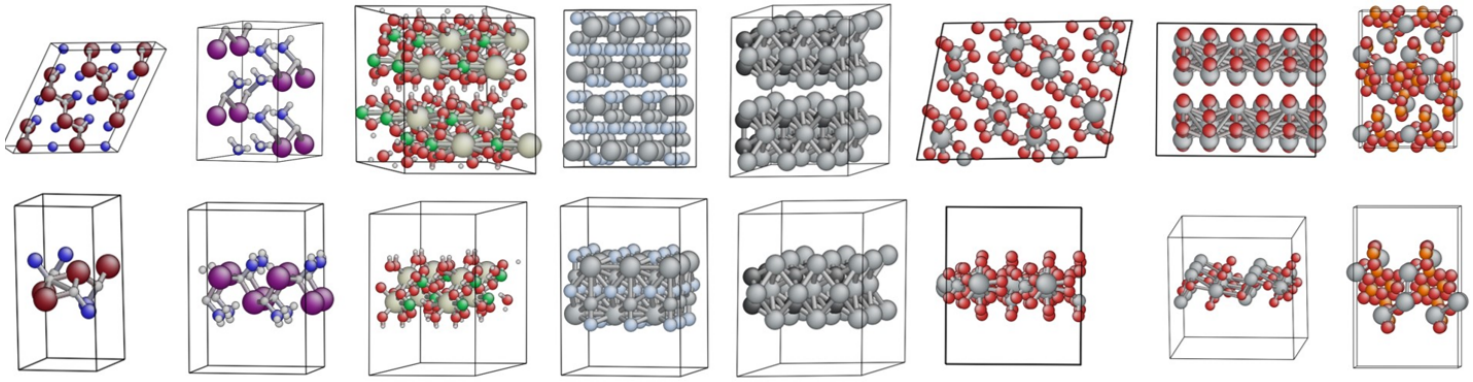
STARTING FROM ICSD/COD DATABASE:

- **108 423** unique 3D structures
- **5619** layered structures
- >**100 000** DFT calculations
- >**30 000** material properties
- >**1·10^{9}** attributes

N. Mounet, GP et al., arXiv:1611.05234 (2016)

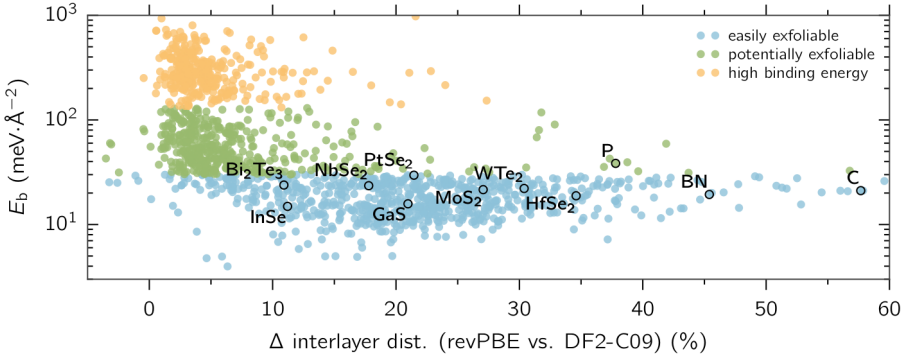
Importance of provenance: an high-throughput example

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
- >$\mathbf{1 \cdot 10^9}$ attributes

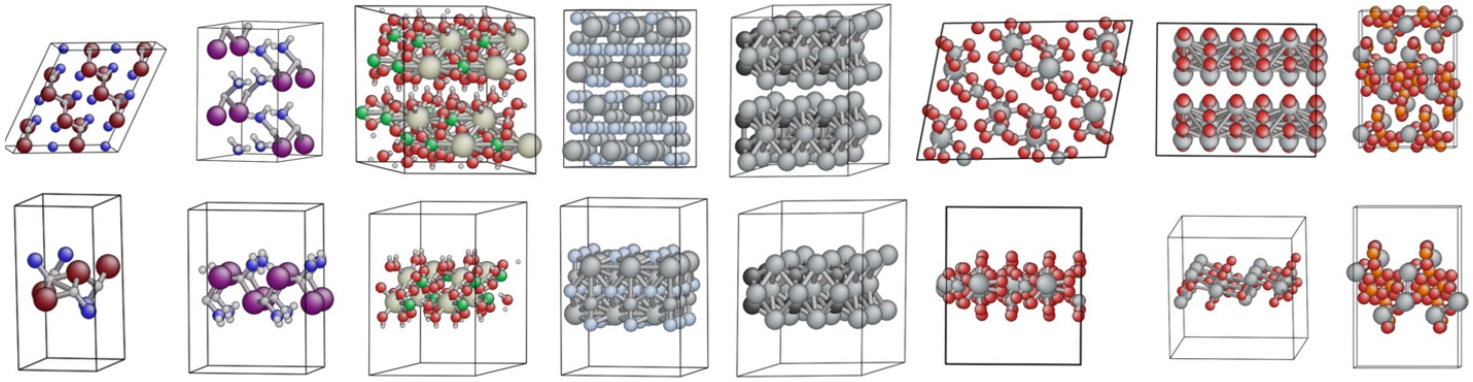


All data needs to be condensed in a few plots

N. Mounet, GP et al., arXiv:1611.05234 (2016)

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
- >**1·10^{9}** attributes

III. METHODS

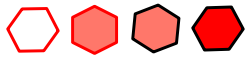
A. Reproducibility and provenance

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.

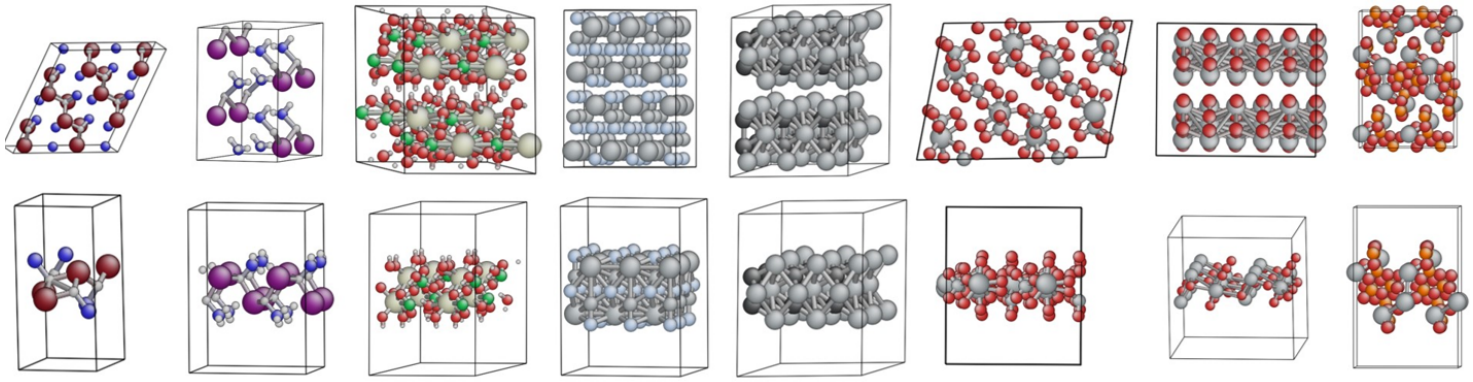
Impossible to describe every detail in the Methods

N. Mounet, GP et al., arXiv:1611.05234 (2016)



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
- >$\mathbf{1 \cdot 10^9}$ attributes

III. METHODS

A. Reproducibility and provenance

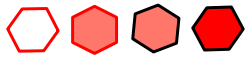
It is an often repeated tenet that results of scientific research must be reproducible. This objective is, however, challenging, especially in high-throughput research, due to the large number of simulations involved and the complex sequence of logical steps needed in the study. To ensure reproducibility 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

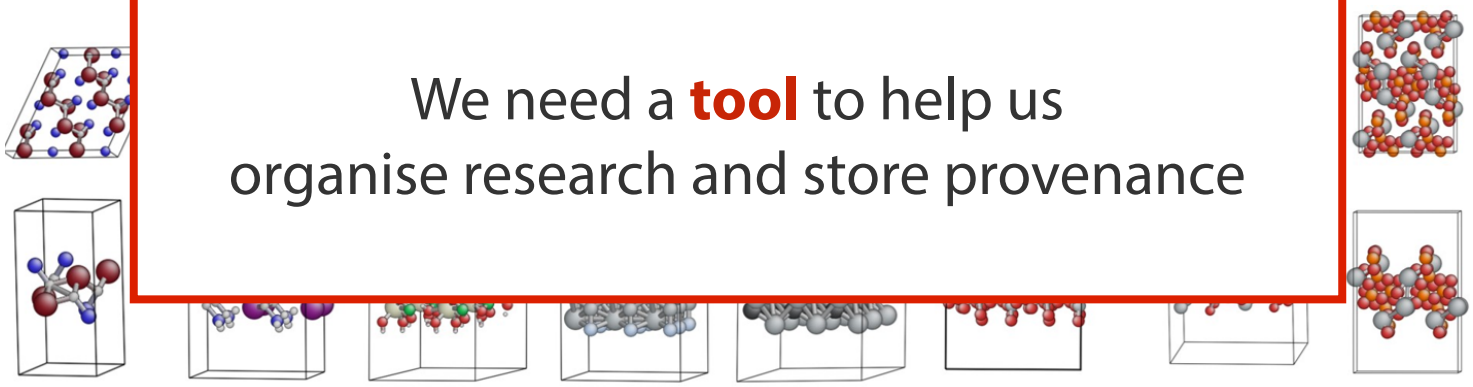
*For **peers**, reproducing data is challenging if not impossible*
*But even for **authors**, it is extremely difficult*

N. Mounet, GP et al., arXiv:1611.05234 (2016)



Importance of provenance: an high-throughput example

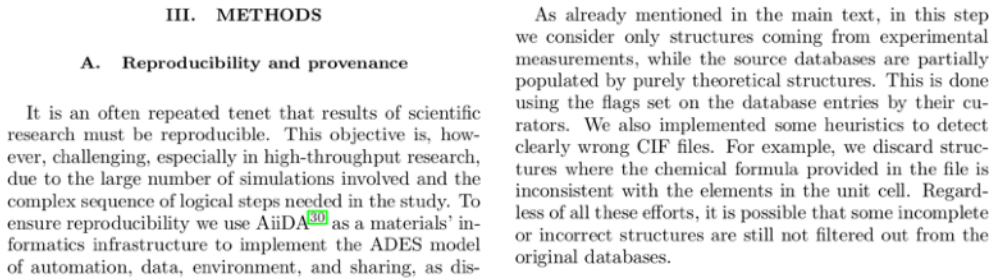
DISCOVERING NEW TWO-DIMENSIONAL MATERIALS



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

STARTING FROM ICSD/COD DATABASE:

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

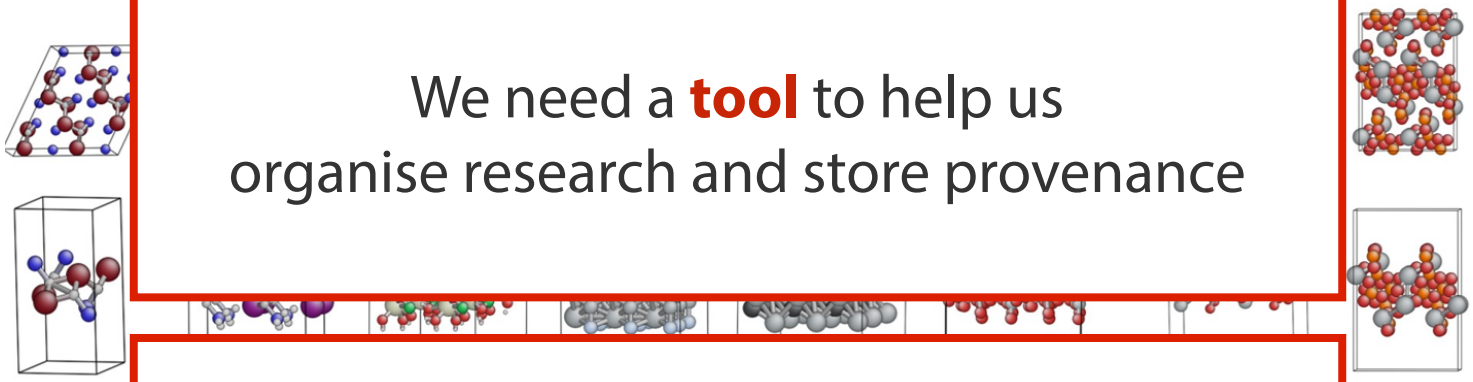
Impossible to describe every detail in the Methods

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

N. Mounet, GP et al., arXiv:1611.05234 (2016)

Importance of provenance: an high-throughput example

DISCOVERING NEW TWO-DIMENSIONAL MATERIALS



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

AiiDA

Automated Interactive Infrastructure and
Database for Computational Science

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

STARTING FRO

- **108 423** uni
- **5619** layere
- >**100 000** D
- >**30 000** ma
- >**1·10^9** attrib

oned in the main text, in this step
ructures coming from experimental
the source databases are partially
theoretical structures. This is done
n the database entries by their cu-
lemented some heuristics to detect
les. For example, we discard struc-
nical formula provided in the file is
elements in the unit cell. Regard-
s, it is possible that some incomplete
es are still not filtered out from the

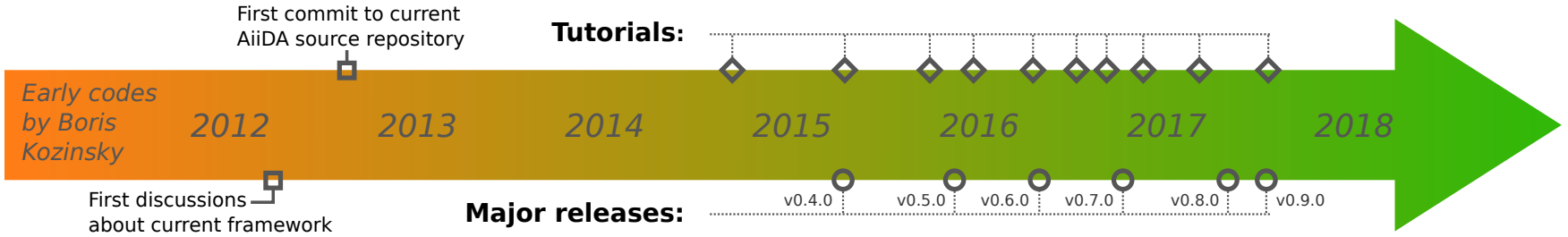
in the Methods

N. Mounet, GP et al., arXiv:1611.05234 (2016)

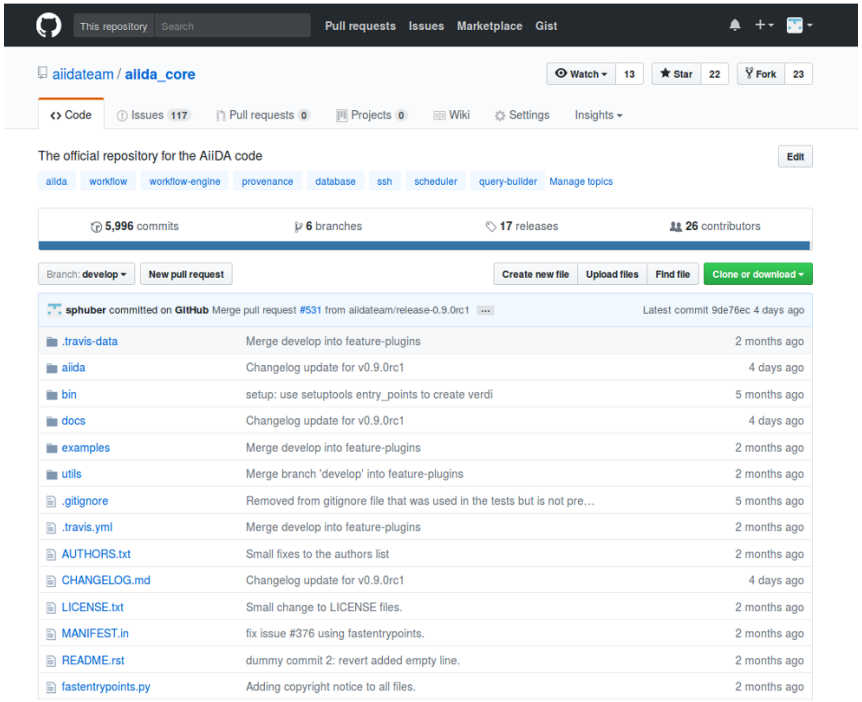
AiiDA

Automated Interactive Infrastructure and Database for Computational Science

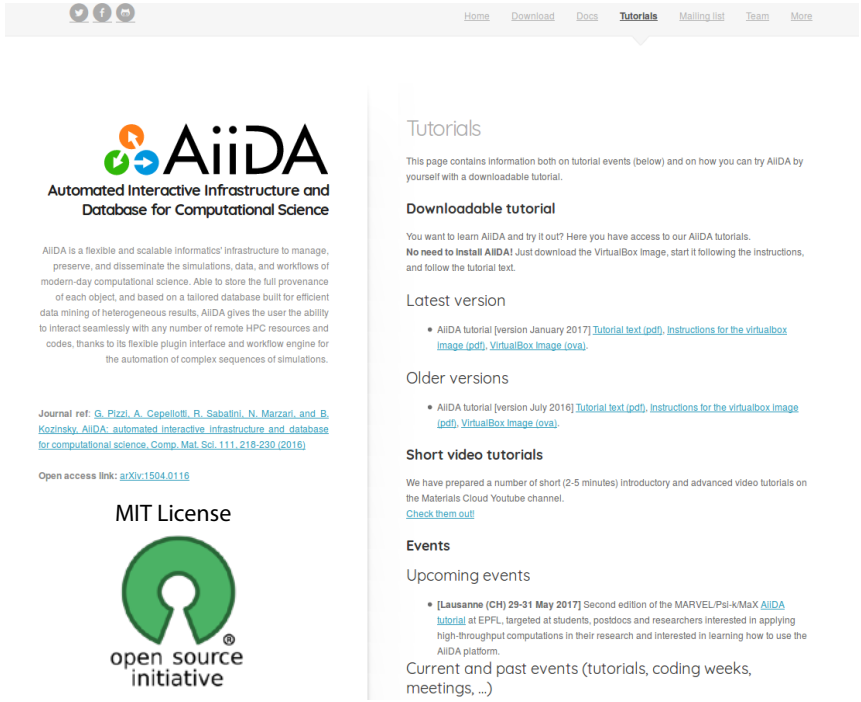
The history of the code



https://github.com/aiidateam/aiida_core



http://aiida.net



Such tool should help make computational research:

reproducible

Often not possible from the data reported in papers

searchable

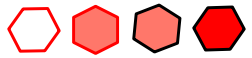
Find existing calculations, reuse and data-mine results

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

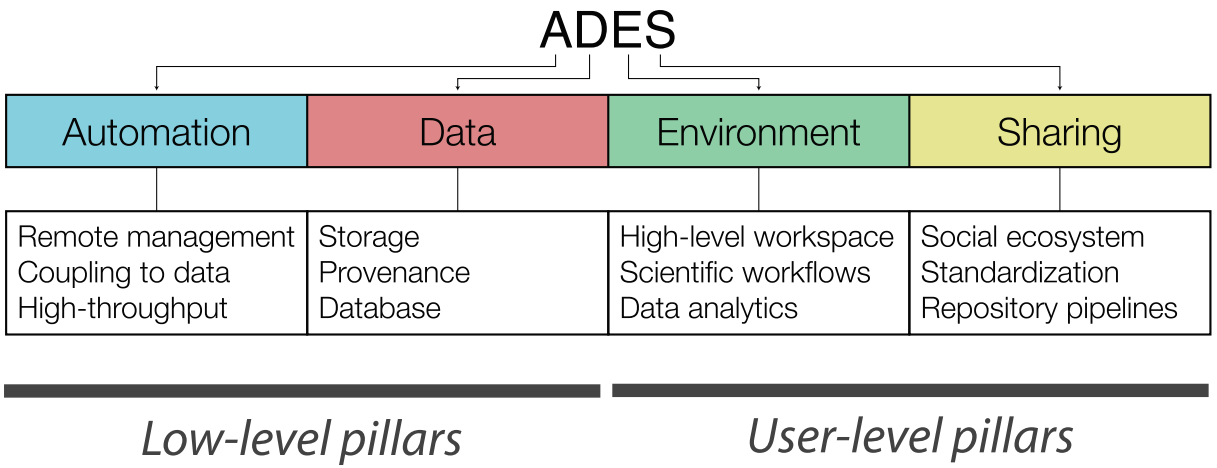
reliable

Automated workflows to verify results and reduce the chance of errors

shareable

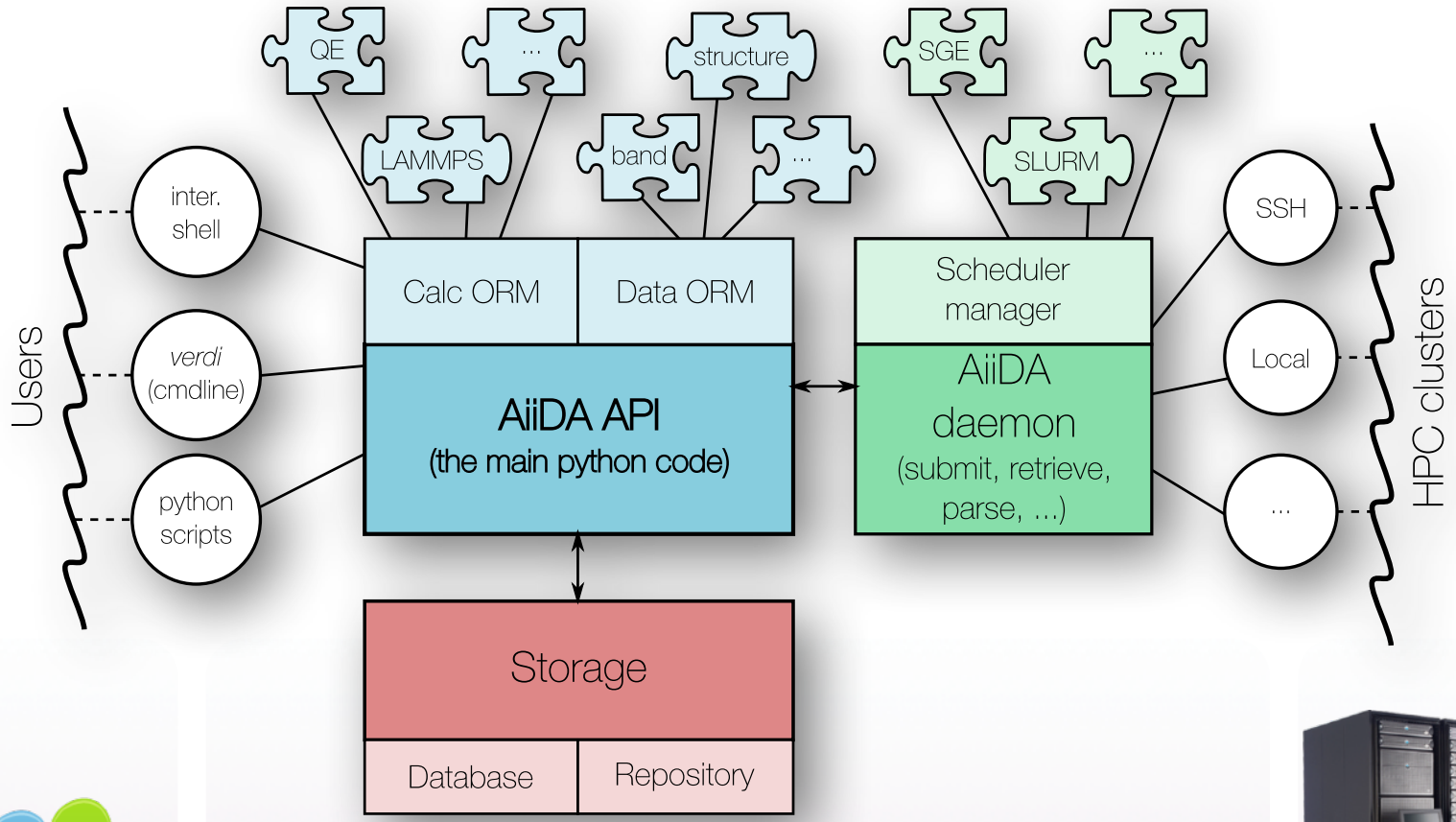
Community to share results, cross-validate them, and boost scientific discovery

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

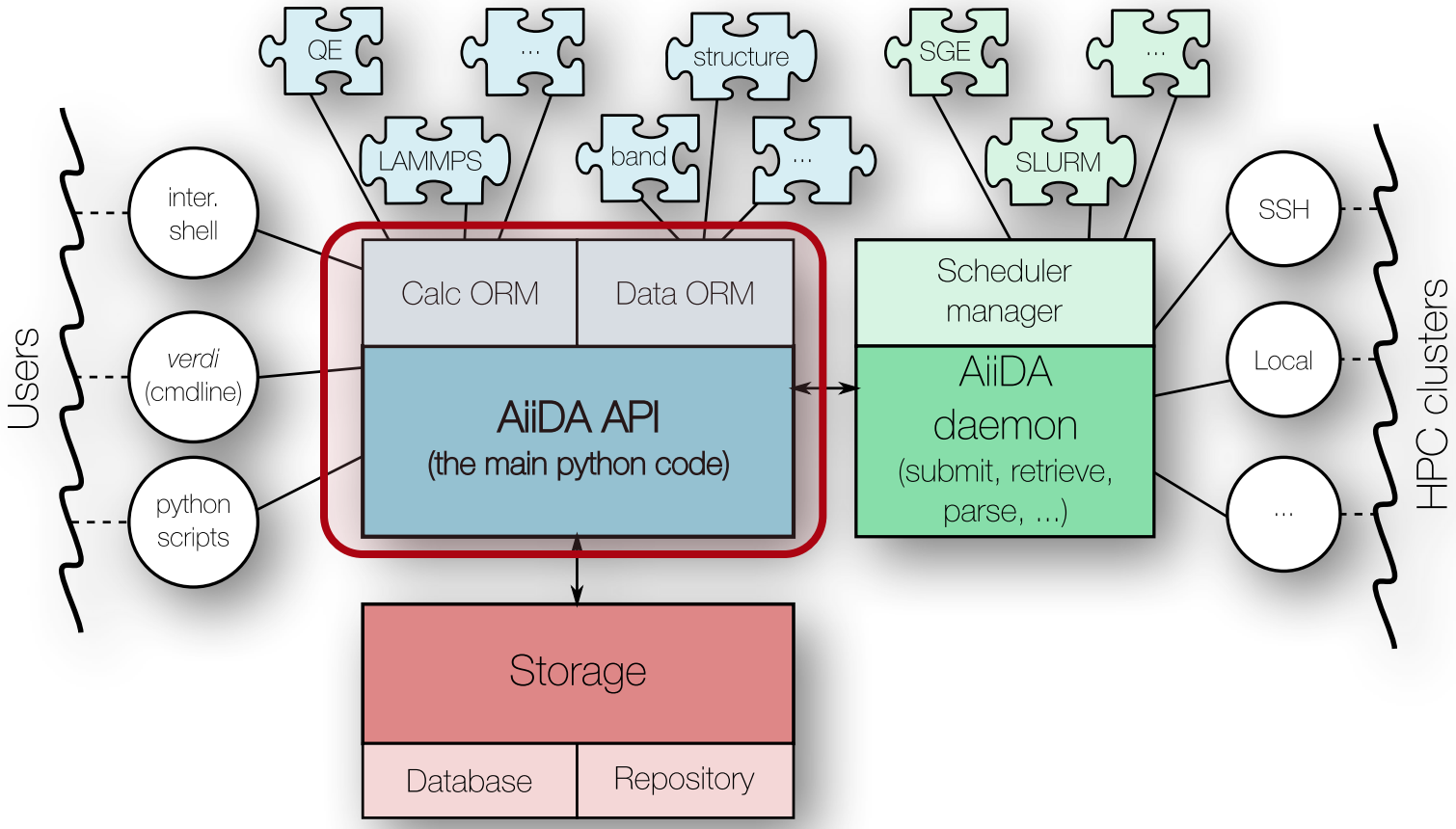


G. Pizzi et al., Comp. Mat. Sci. 111, 218-230 (2016)

What is AiiDA?

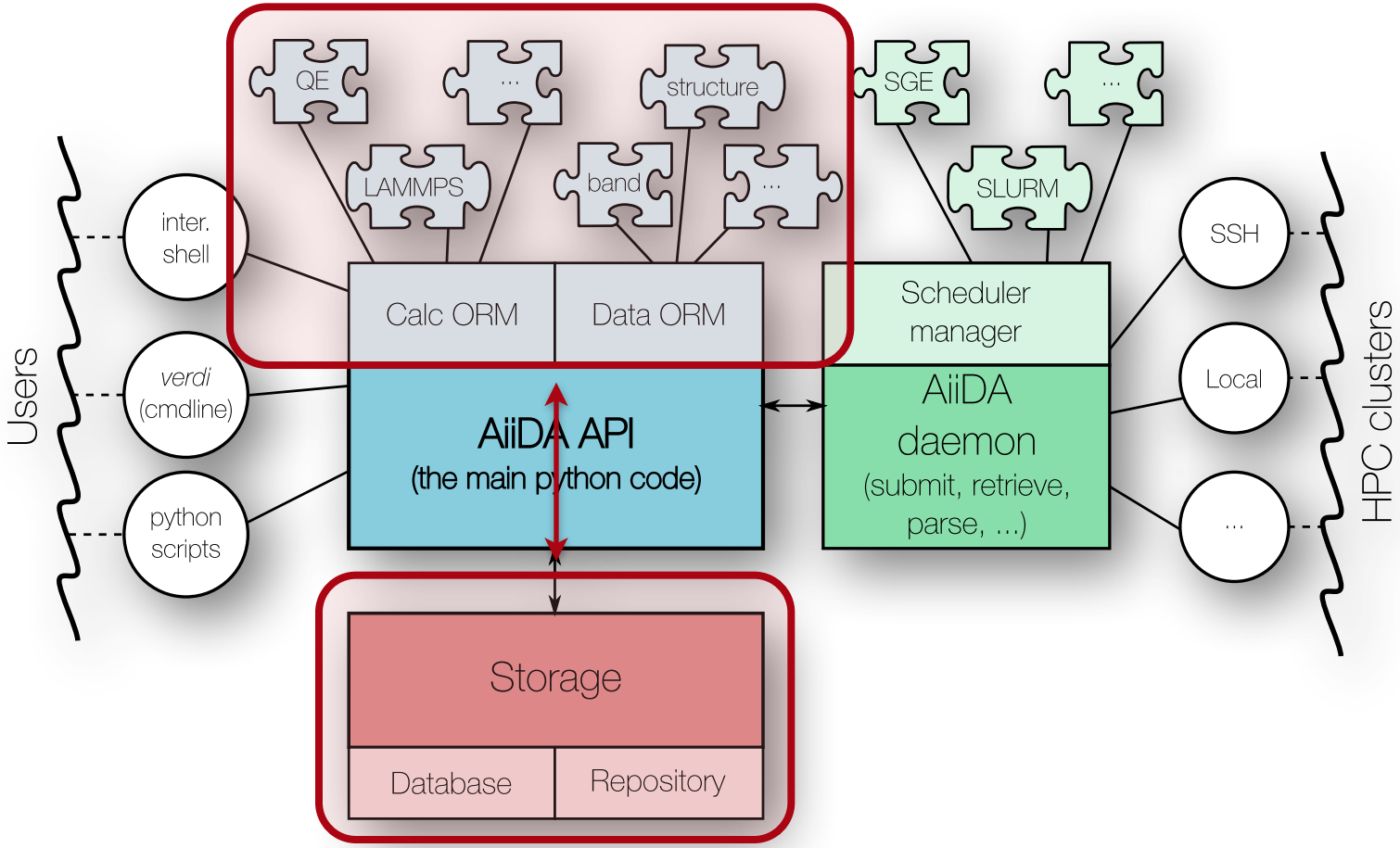


What is AiiDA?



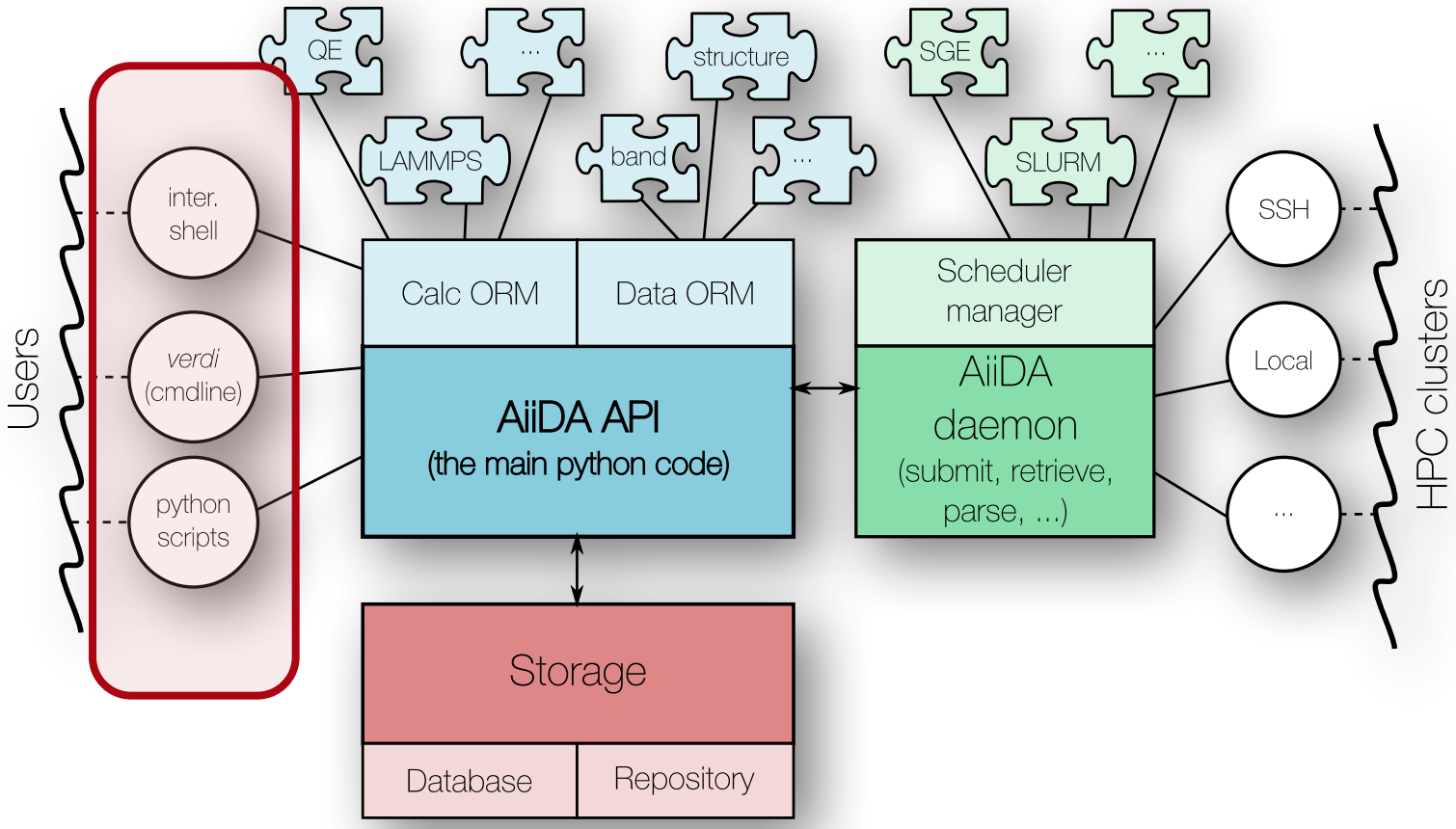
1. The core of the code is the **AiiDA API:** the main python code and classes

What is AiiDA?



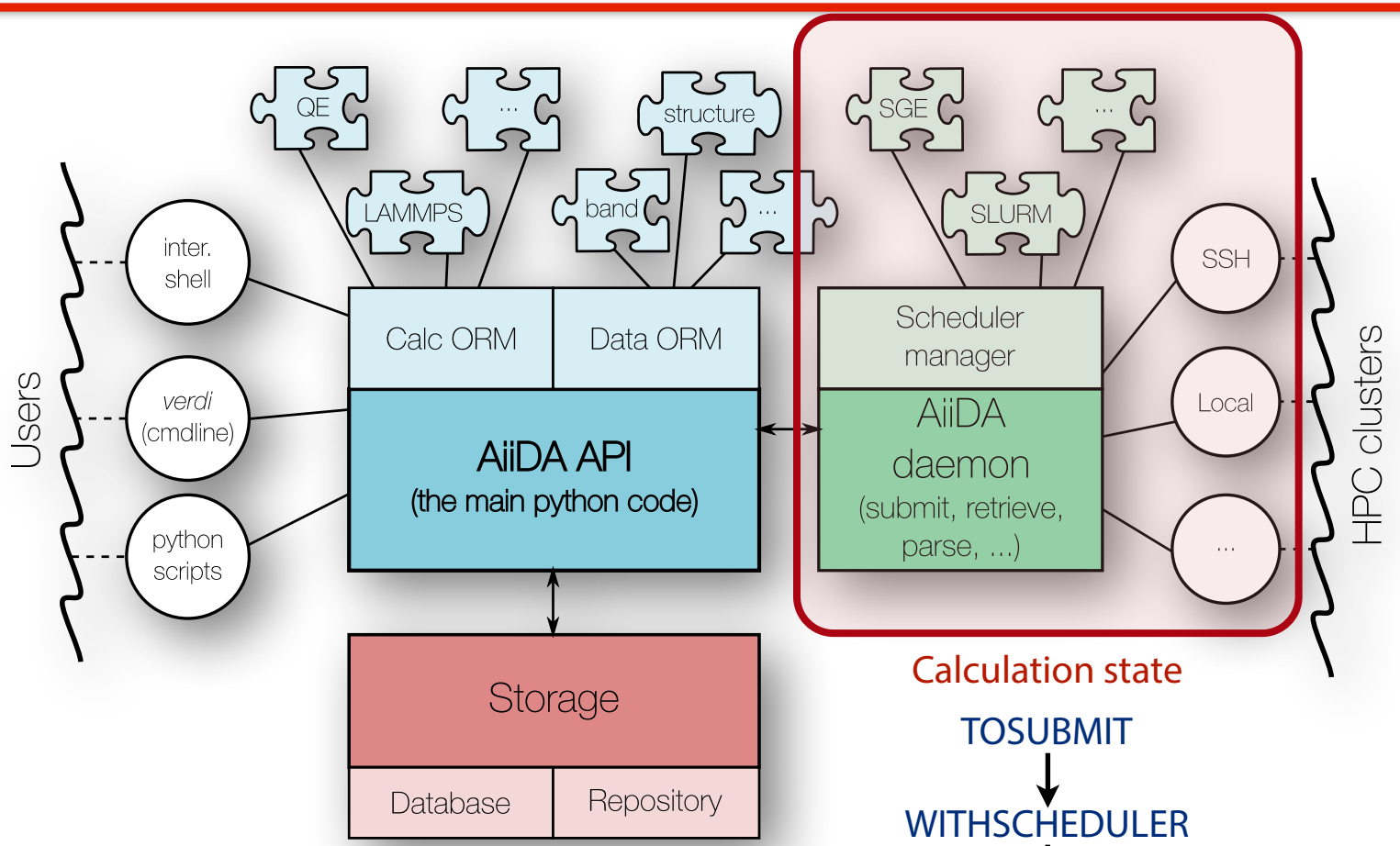
2. The **AiiDA Object-Relational Mapper** (ORM): transparently store data, codes and calculations in a database, *transparent to the user*

What is AiiDA?



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

What is AiiDA?



Calculation state

TOSUBMIT

↓

WITHSCHEDULER

↓

RETRIEVED

↓

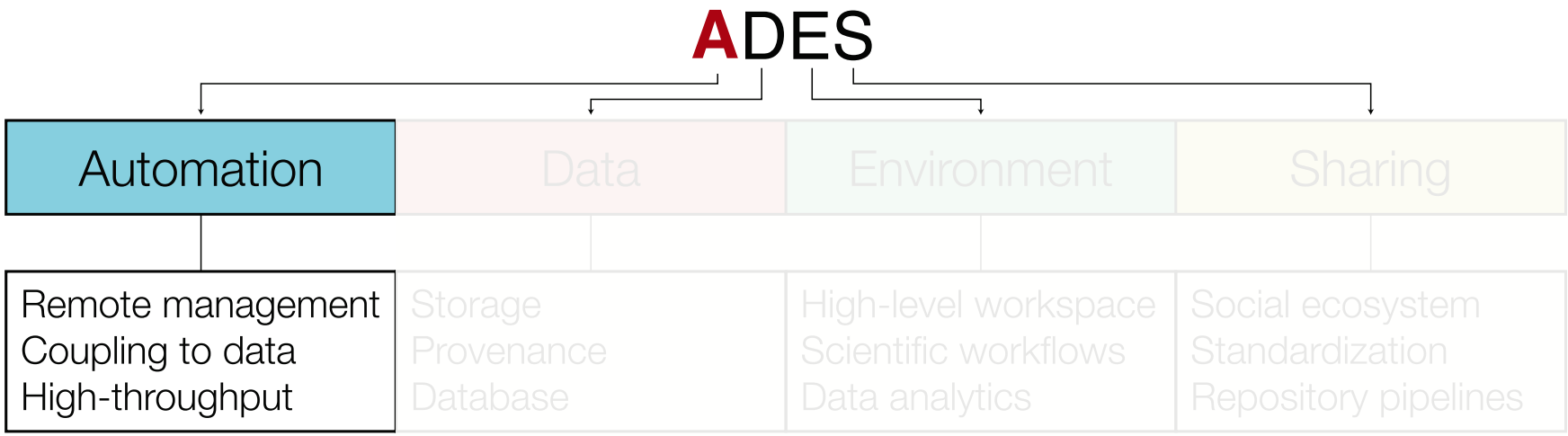
PARSED

↓

FINISHED

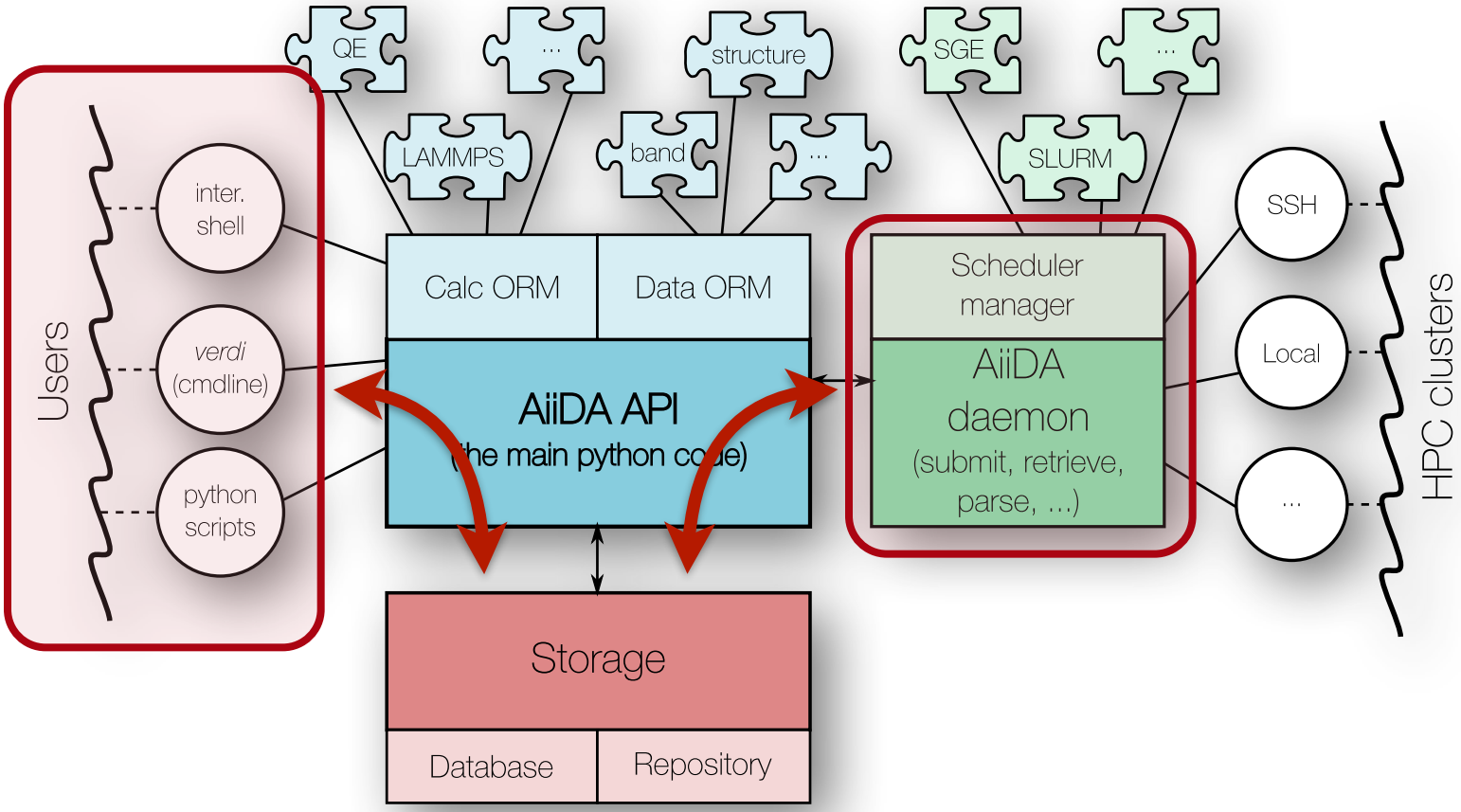
4. A **daemon** to manage interaction with remote computers without user intervention

Automation in AiiDA



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)

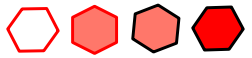
Abstraction into APIs: a single calculation

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

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

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
```

Choose code and computer

Define all inputs

Take care of running the calculation
through the daemon



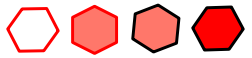
Abstraction into APIs: a single calculation

```
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
```

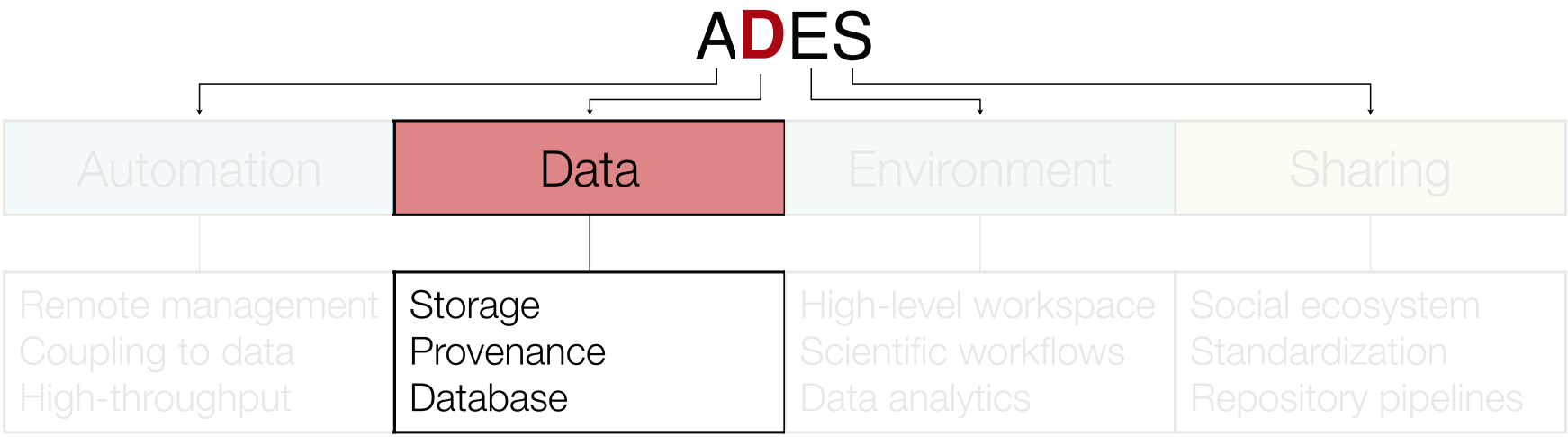
Just one line change to change computer (with different cluster, schedulers, version of codes, …)

Data gets stored in the DB during submission

Take care of running the calculation **through the daemon**



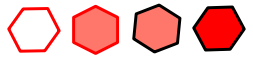
Data in AiiDA



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?**



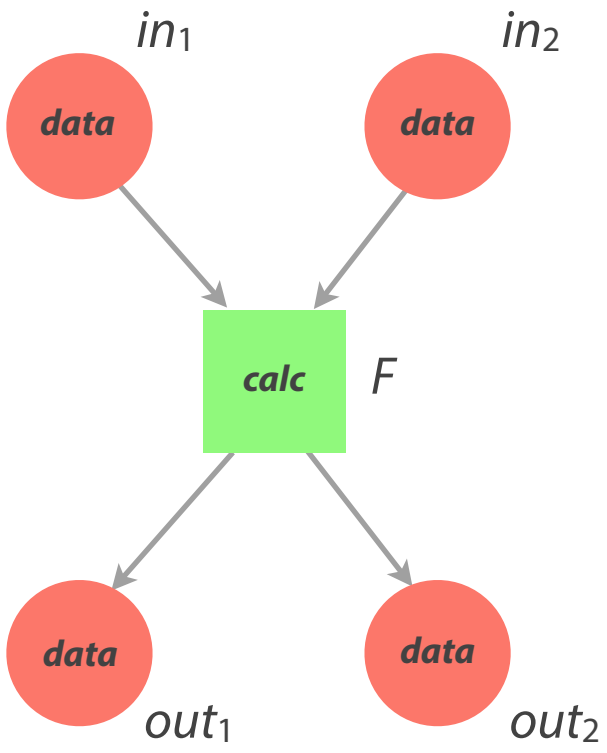
Storage and provenance

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

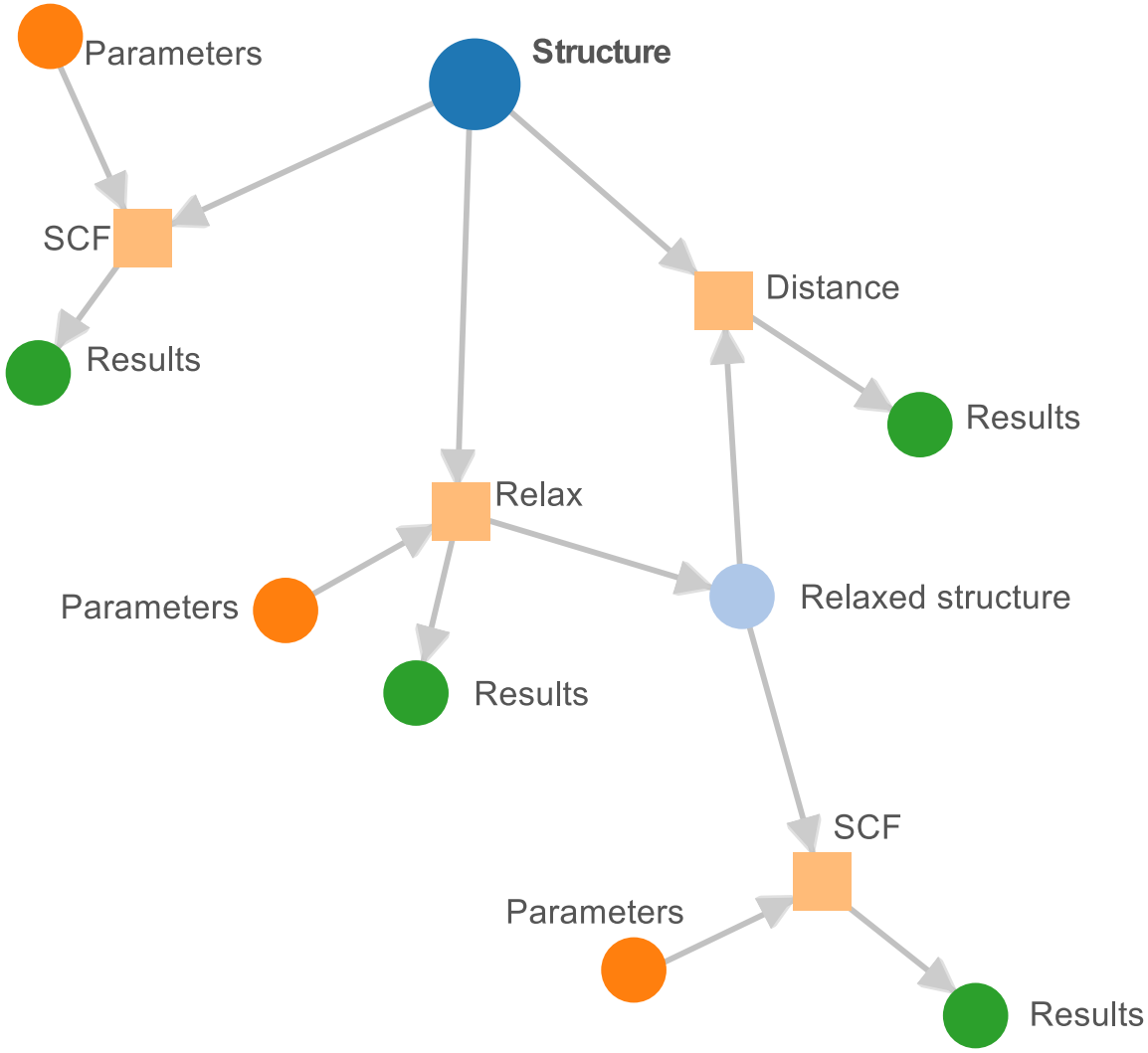
- Inspiration from the *open provenance model*
- **Any calculation: a function,** converting inputs to outputs:

$$out_1, out_2 = \mathbf{F}(in_1, in_2)$$

- **Each object is a node in a graph**, connected by directional labeled links
- Output nodes can be used as inputs

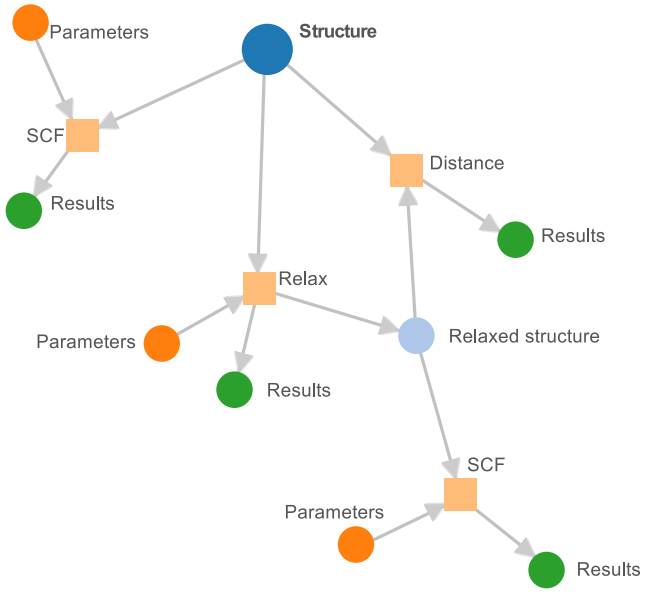


Data provenance: Directed Acyclic Graphs



Data provenance: Directed Acyclic Graphs

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



Data provenance: Directed Acyclic Graphs

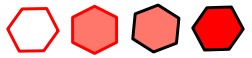
Directed Acyclic Graphs:
appropriate representation
of calculations, data
and their **provenance**

Next questions

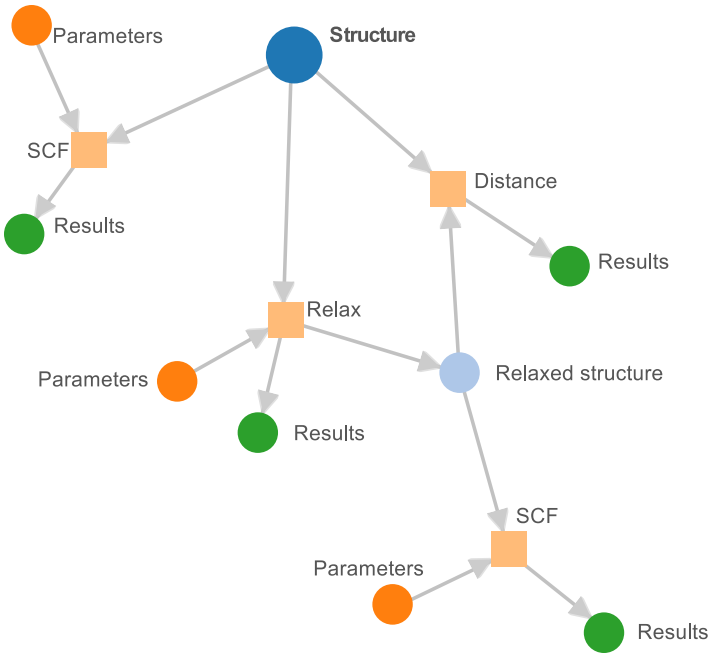
What must we store?

How can we store it?

How can we query it?



Saving the DAGs: Nodes and Links

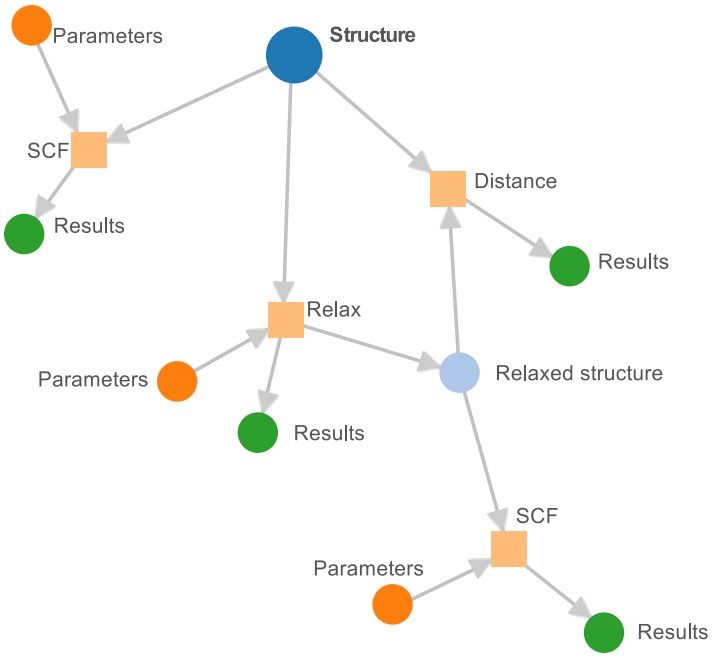


How to represent it… in SQL?

- *Each **node***: row in a SQL table
  - Additional data:
    - key-value **attributes**
    - **Files/folders**

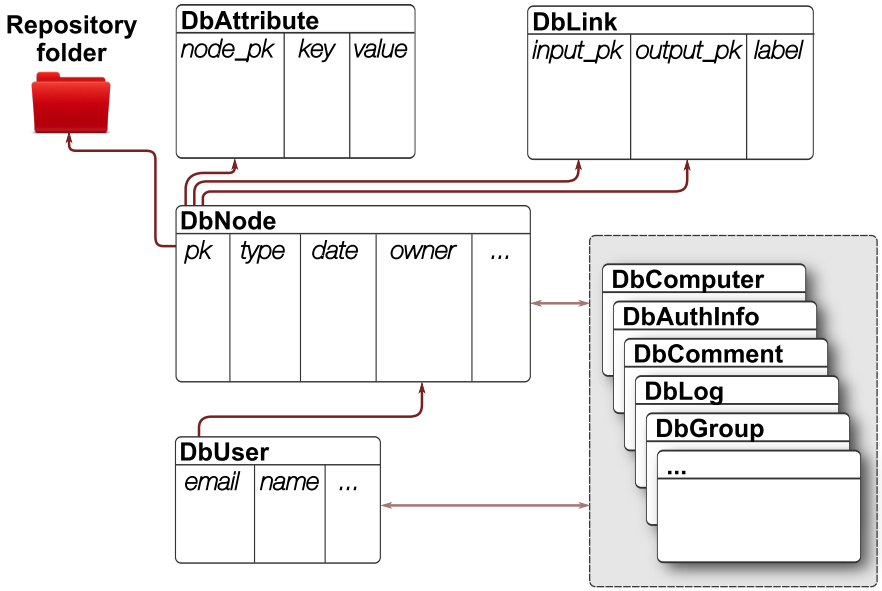
- **Links** also stored in a SQL table
  ⇒ *jobs provenance*

Saving the DAGs: Nodes and Links



How to represent it… in SQL?

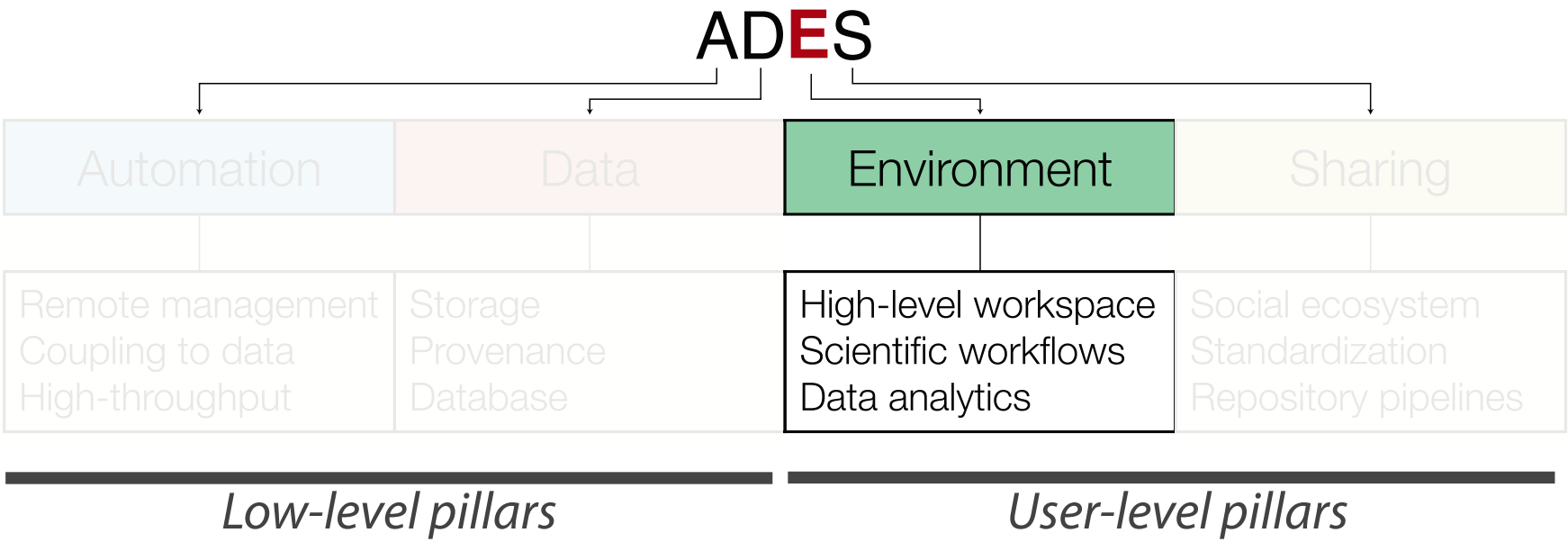
- *Each **node***: row in a SQL table
  - Additional data:
    - key-value **attributes**
    - **Files/folders**
- **Links** also stored in a SQL table
  ⇒ *jobs provenance*



- 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 g*raph DBs* like Neo4j and TitanDB)

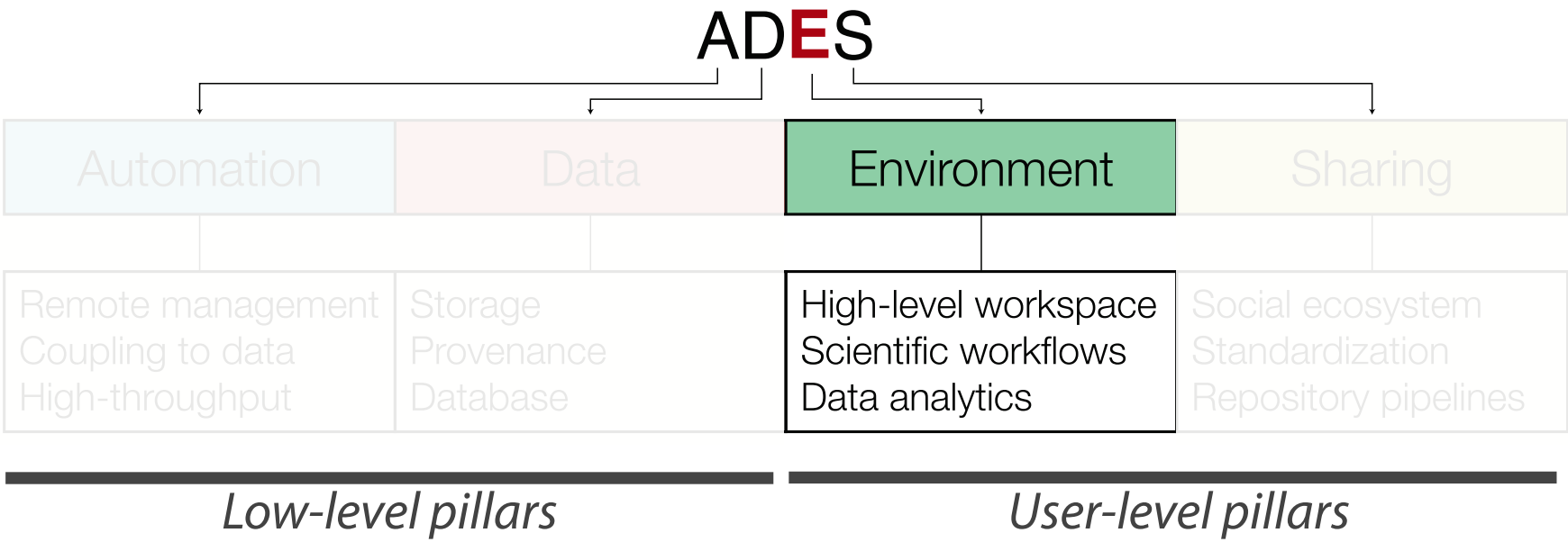
G. Pizzi et al., Comp. Mat. Sci 111, 218-230 (2016)

Environment in AiiDA



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

Environment in AiiDA



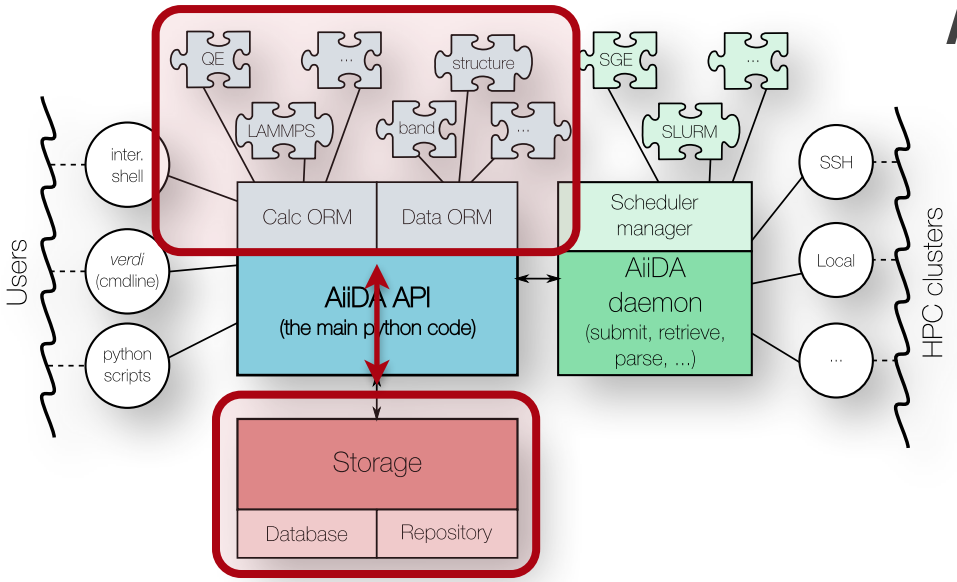
Is AiiDA easy to use?

Does it provide an advantage over files and custom-made scripts?

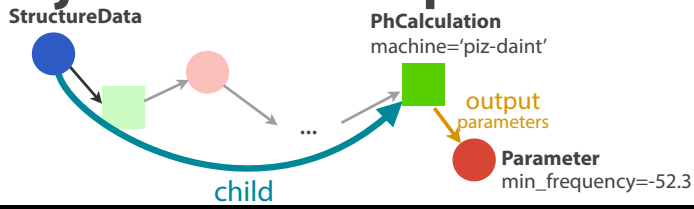
G. Pizzi et al., Comp. Mat. Sci 111, 218-230 (2016)

Environment in AiiDA: user interaction

High level Python interface

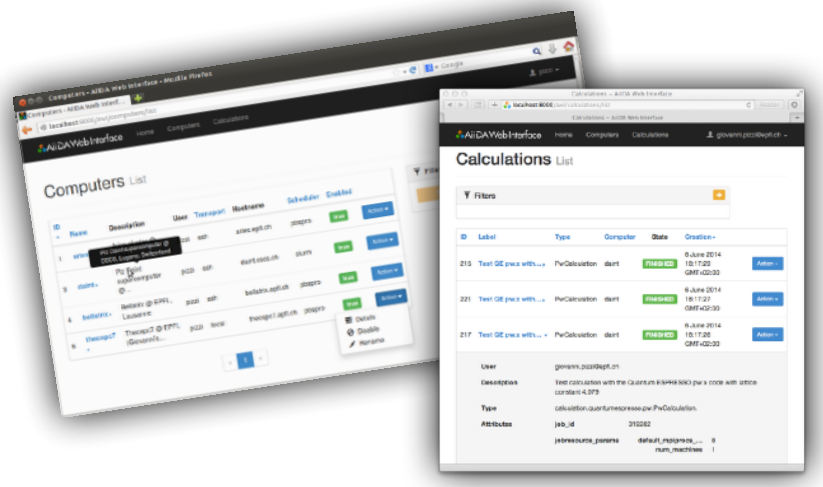


DB-agnostic query interface for AiiDA objects and their provenance




```
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()
```

Seamless user interaction



Extensive documentation + tutorials

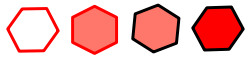
http://aiida-core.readthedocs.org/

Reusability and modularity: AiiDA plugins



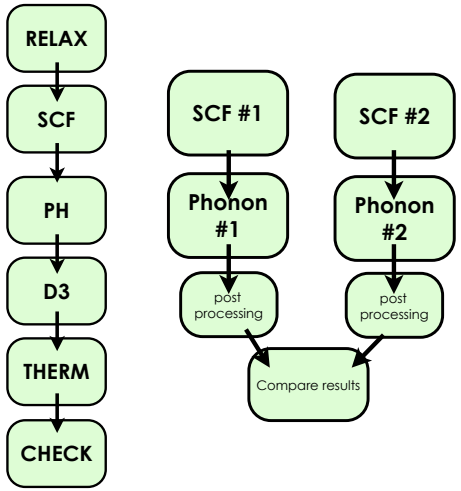


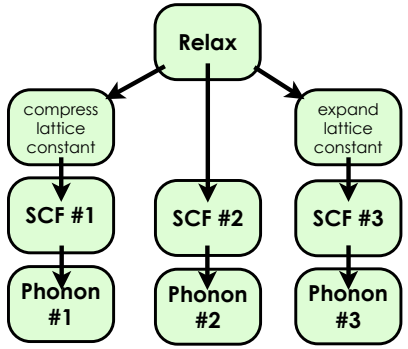

Calculation	**Data**	**Parser**	**Transport**	**Scheduler**	**Importers & exporters**
Generation of input files for a given code	Management of data objects for input/ output	Parsing of code output & generation of new DB nodes	How to connect to a cluster	How to interact with the scheduler	Import structures, ... from external DBs
Quantum ESPRESSO, Phonopy, ASE, GPAW, Yambo, NWChem, …	*files&folders, parameter sets, remote data, structures, pseudos, ...*	*Quantum ESPRESSO, Phonopy, ASE, GPAW, Yambo, NWChem, ...*	*local connection, ssh, ...*	*PBSPro, Torque, SGE, SLURM, LSF, …*	*ICSD, COD, TCOD, MPOD, ...*



Environment in AiiDA: Scientific workflows

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

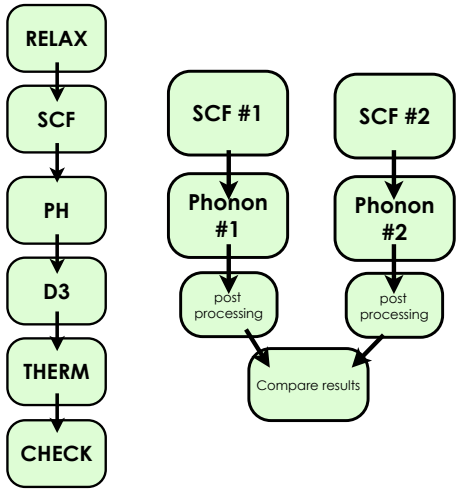


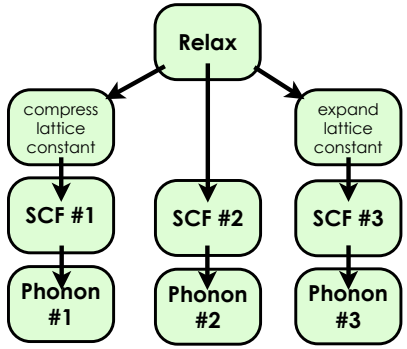


Environment in AiiDA: Scientific workflows

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

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



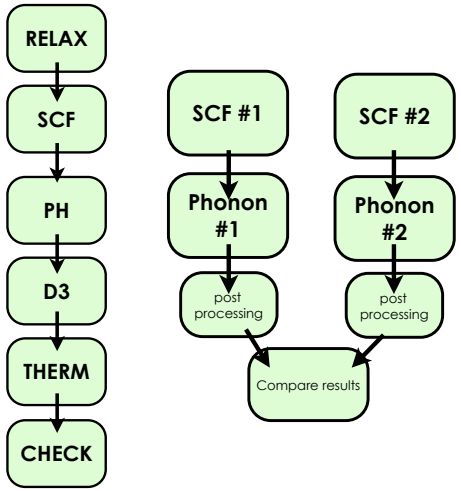


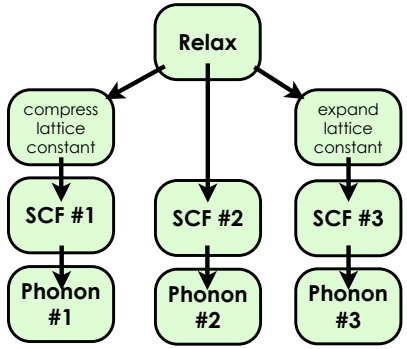
Environment in AiiDA: Scientific workflows

- 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

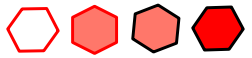




Workflows - a 'Hello World' example

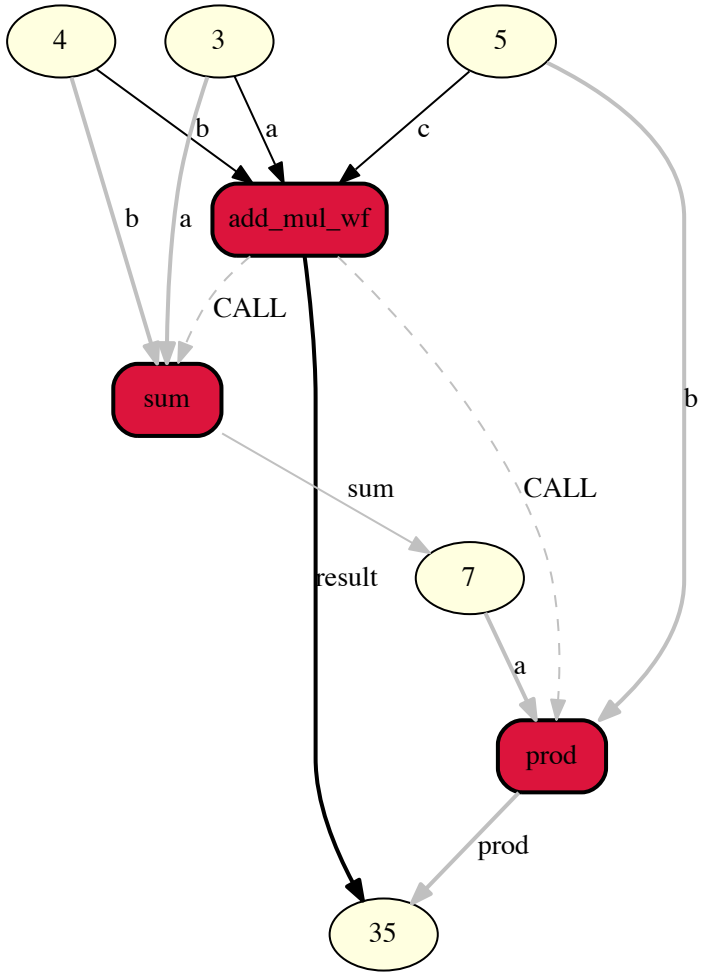
```
@wf
def sum(a, b):
    return {'sum': a+b}


@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']
```

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']
```



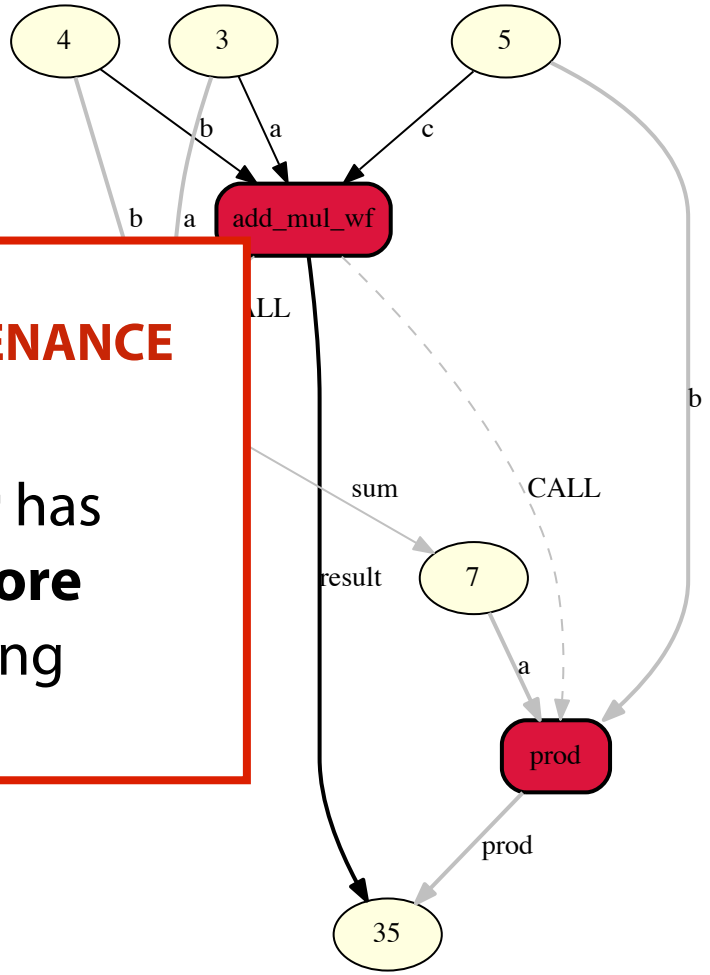
final_value = (3+4) * 5 = 35

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']
```

AUTOMATIC PROVENANCE TRACKING!

The user **never** has to explicitly **store** or **link** anything

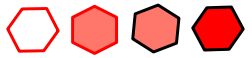


final_value = (3+4) * 5 = 35

AiiDA's workflow engine features

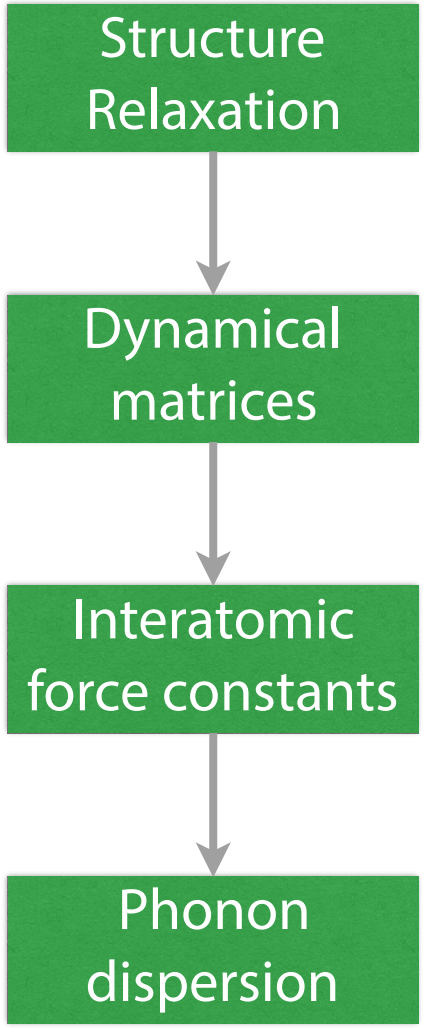
- **Workfunctions** are not a cure-all and have limitations
  - Interpreter is blocked during execution
  - Misses "sequential" character of execution

- 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



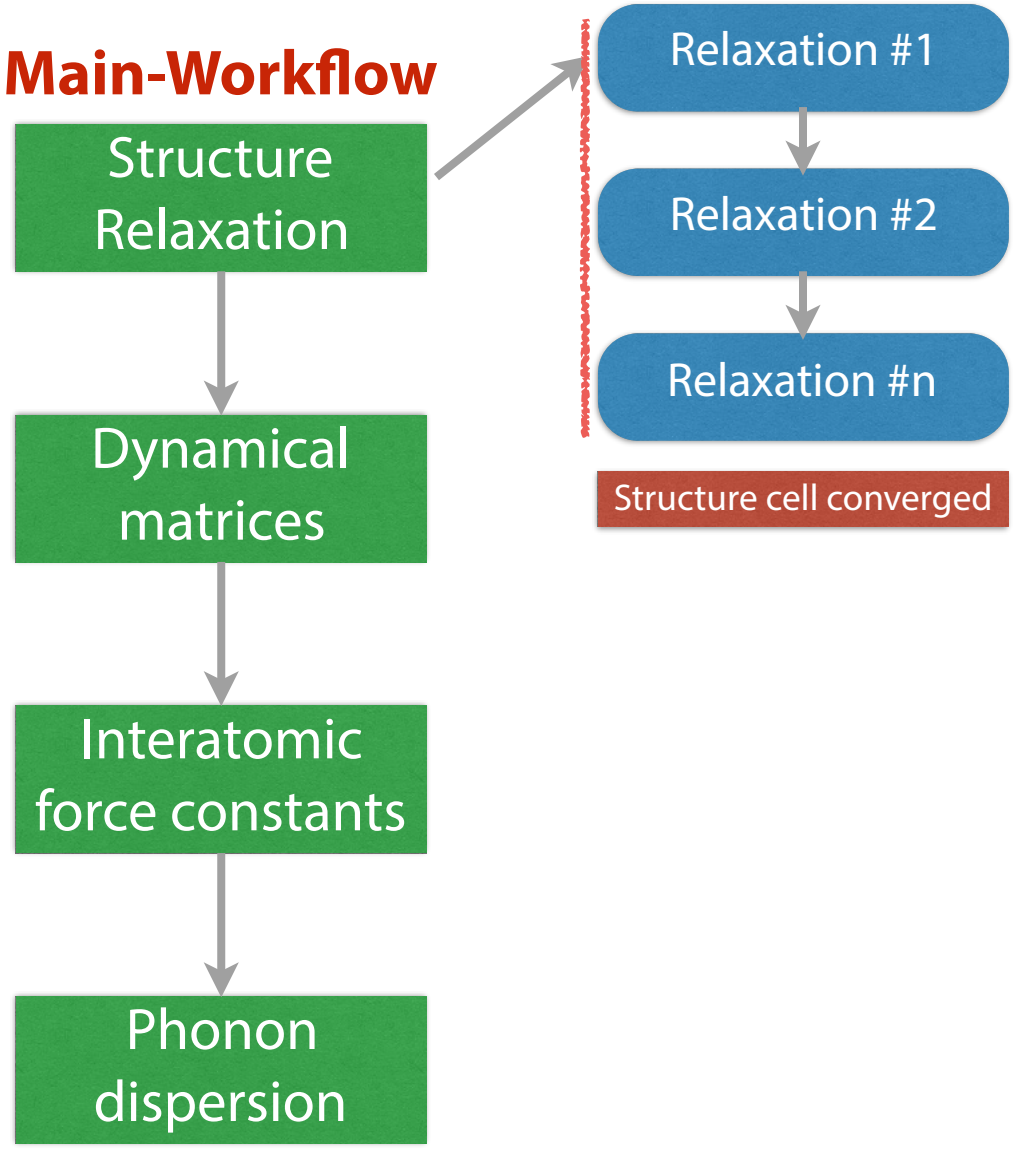
A real-life workflow example: phonon dispersions

Main-Workflow



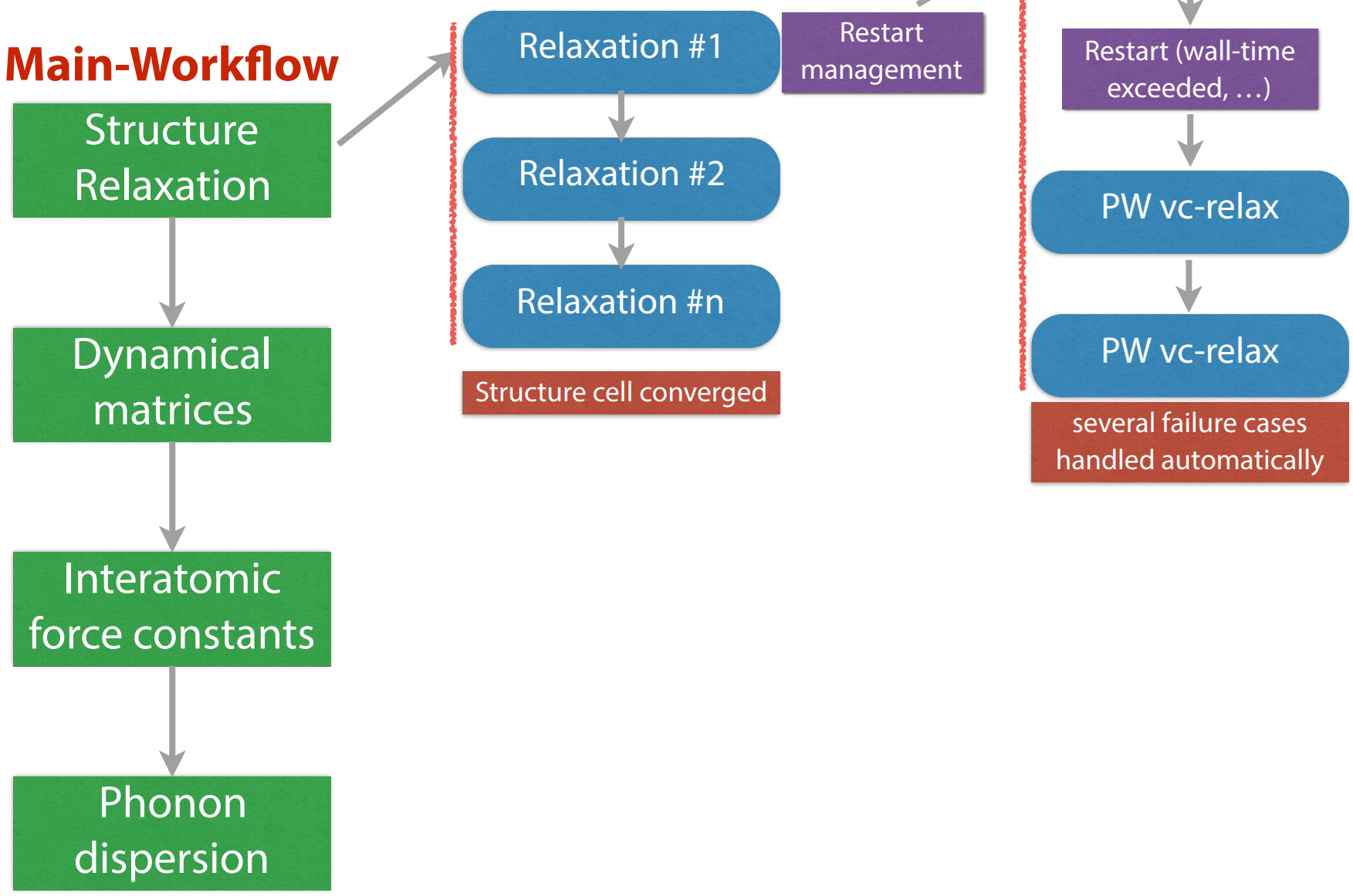
N. Mounet et al.

A real-life workflow example: phonon dispersions



N. Mounet et al.

A real-life workflow example: phonon dispersions



N. Mounet et al.

A real-life workflow example: phonon dispersions

Main-Workflow

Sub-workflows

Structure Relaxation

Dynamical matrices

Single calculations

Interatomic force constants

Phonon dispersion

Relaxation #1

Relaxation #2

Relaxation #n

Structure cell converged

Restart management

PW vc-relax

Restart (wall-time exceeded, …)

PW vc-relax

PW vc-relax

several failure cases handled automatically

Initialize PH

PH on q-grid

Collect phonons

Parallelization

PH on $\mathbf{q}_1$

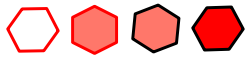
PH on $\mathbf{q}_2$

PH on $\mathbf{q}_n$

N. Mounet et al.

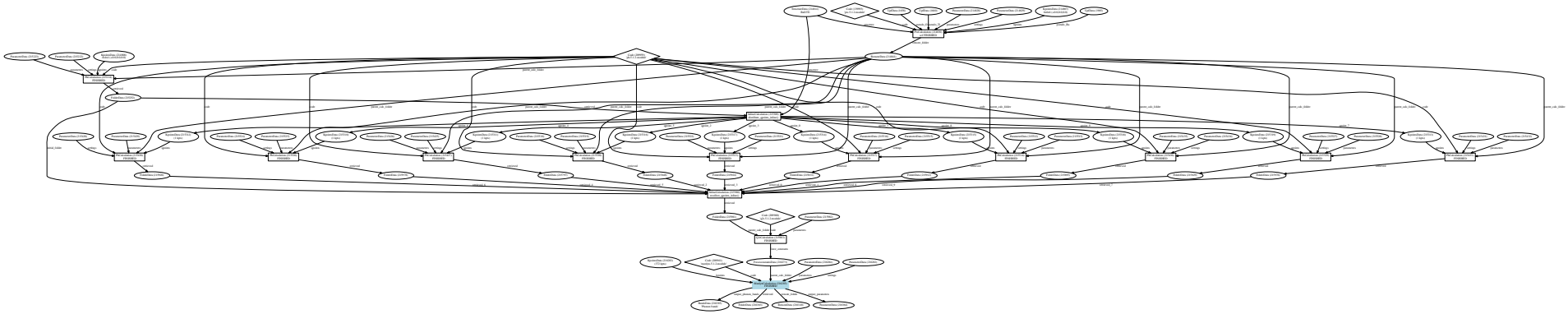
A real workflow

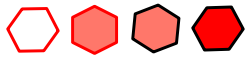
```
params = {'input': {'kpoints_density': 0.2,
                    'convergence': 'tight'},
          'structure': structure,
          'pseudo_family': pseudo_family,
          'machinename': 'mycluster',
          'pw_input':{'volume_conv_threshold': 5e-2},
                      'pw_parameters': {
                         '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

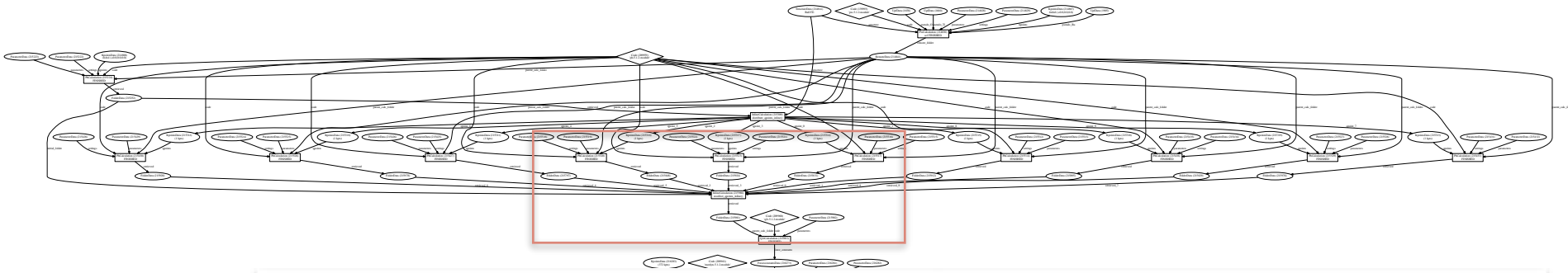
```
params = {'input': {'kpoints_density': 0.2,
                    'convergence': 'tight'},
          'structure': structure,
          'pseudo_family': pseudo_family,
          'machinename': 'mycluster',
          'pw_input':{'volume_conv_threshold': 5e-2},
          'pw_parameters': {
          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



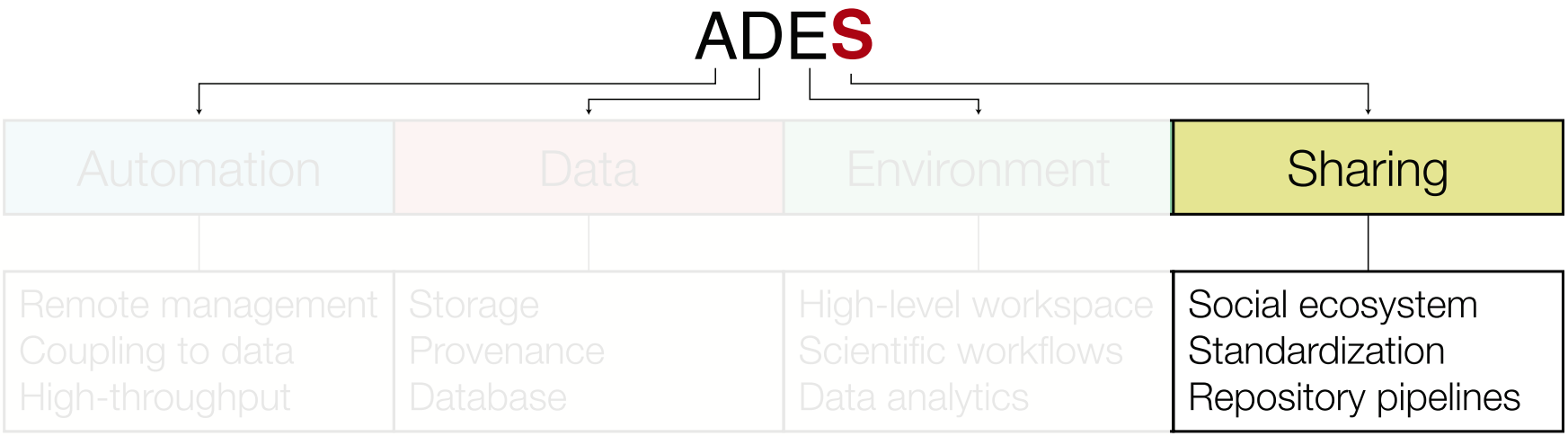


A real workflow

```
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)
```



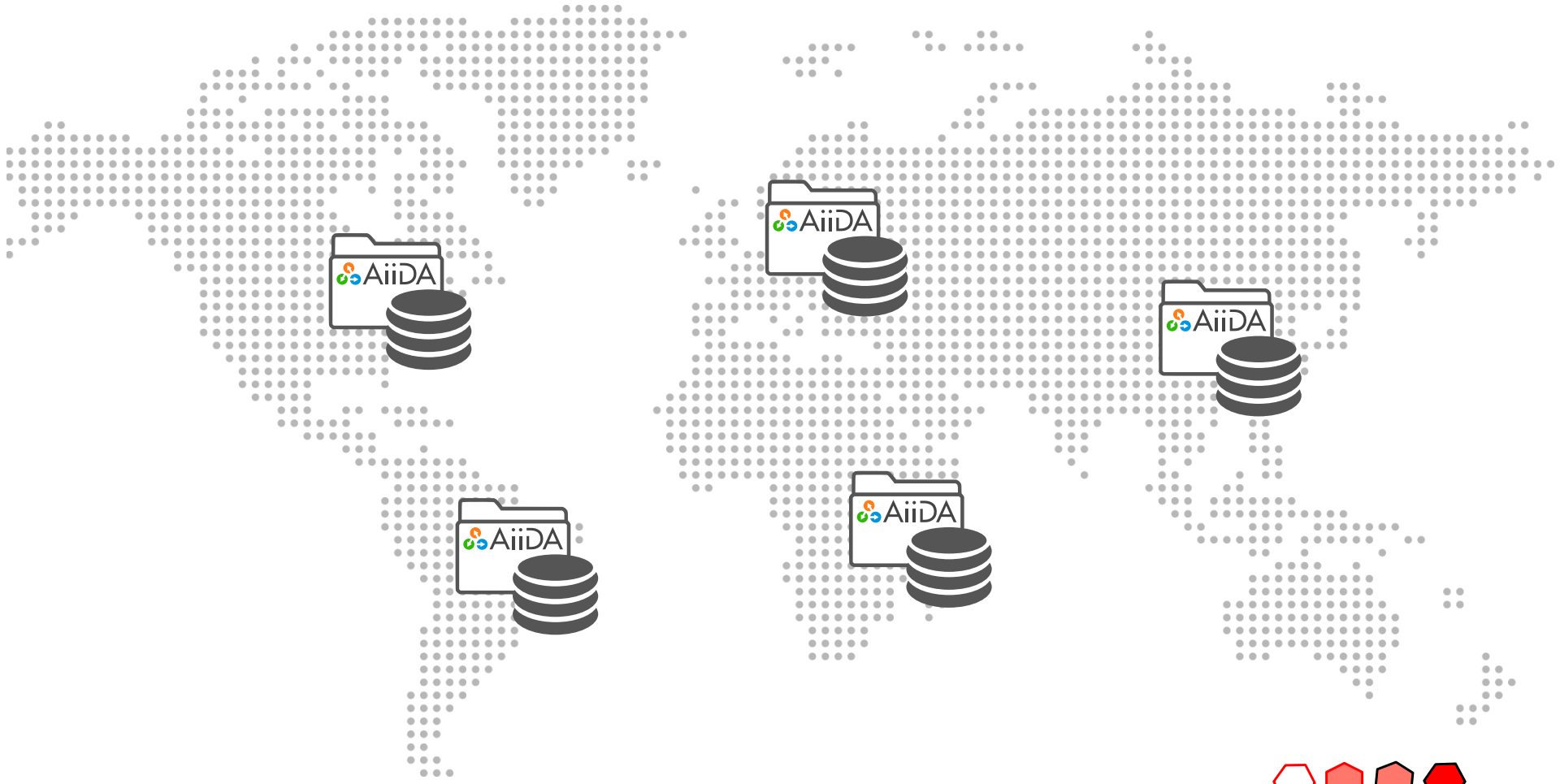
Sharing in AiiDA



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

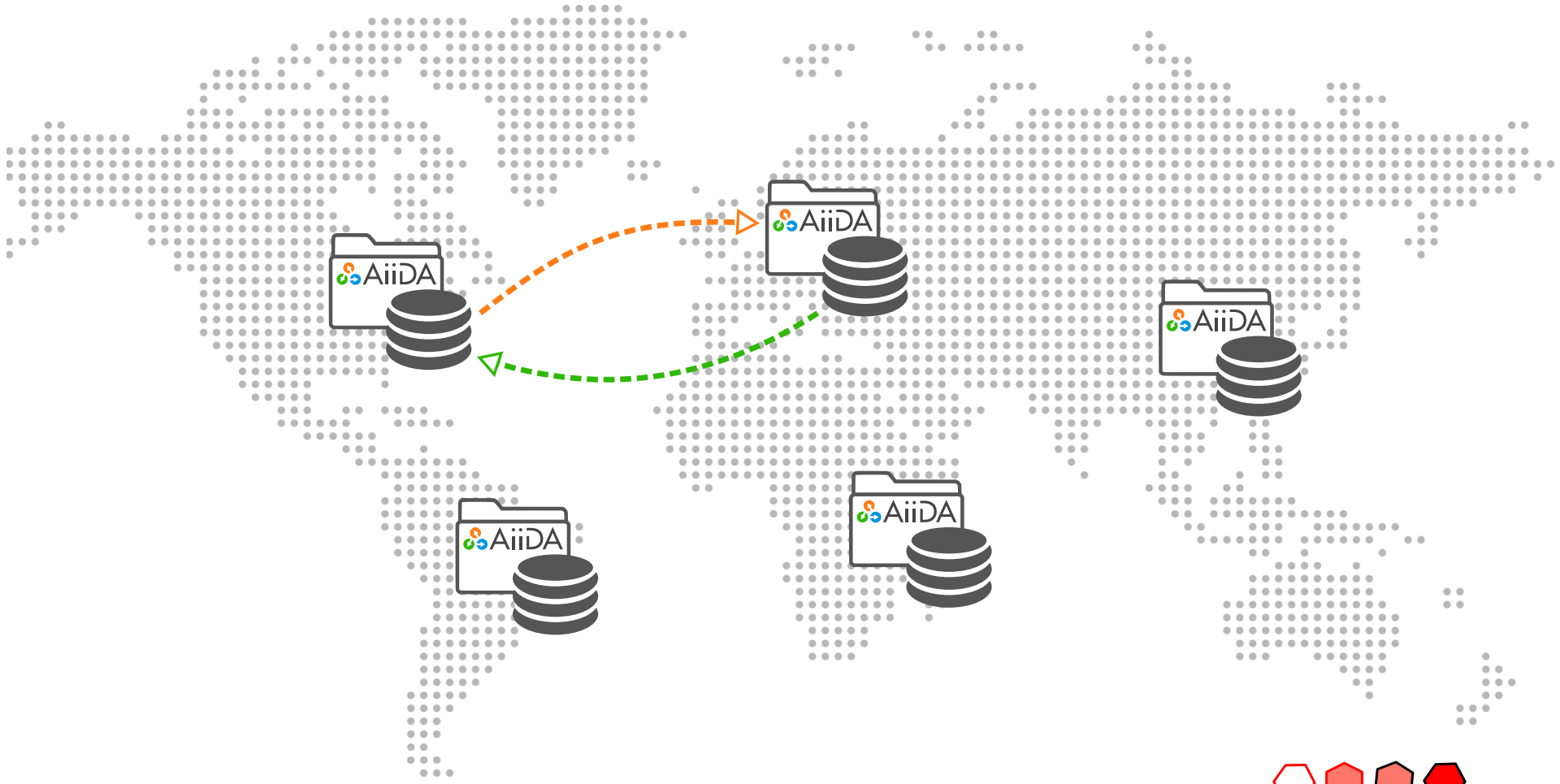
Sharing in AiiDA

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



Sharing in AiiDA

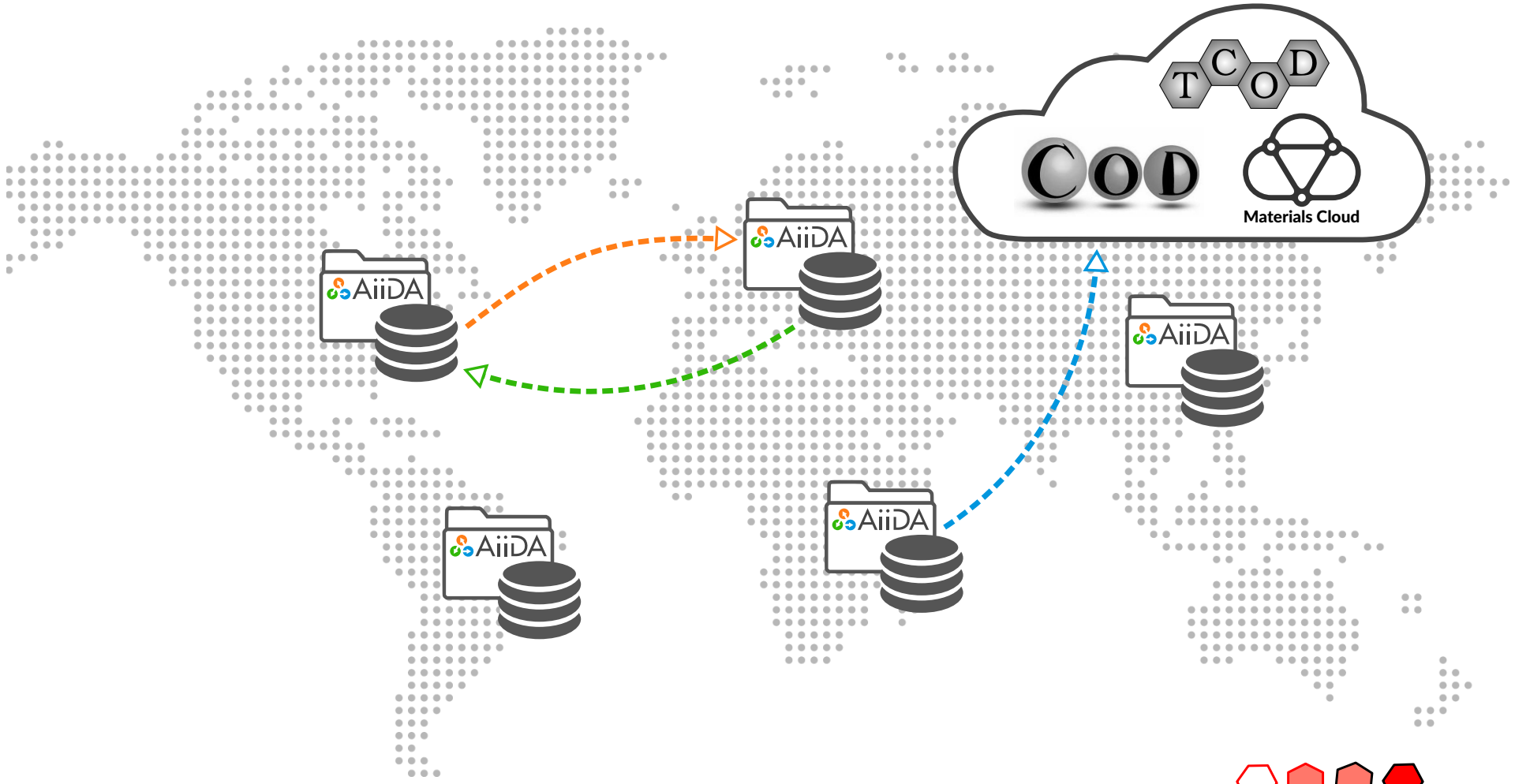
- *Private* AiiDA instances
- UUIDs to uniquely identify nodes
- Data can be shared to other AiiDA repositories



Sharing in AiiDA

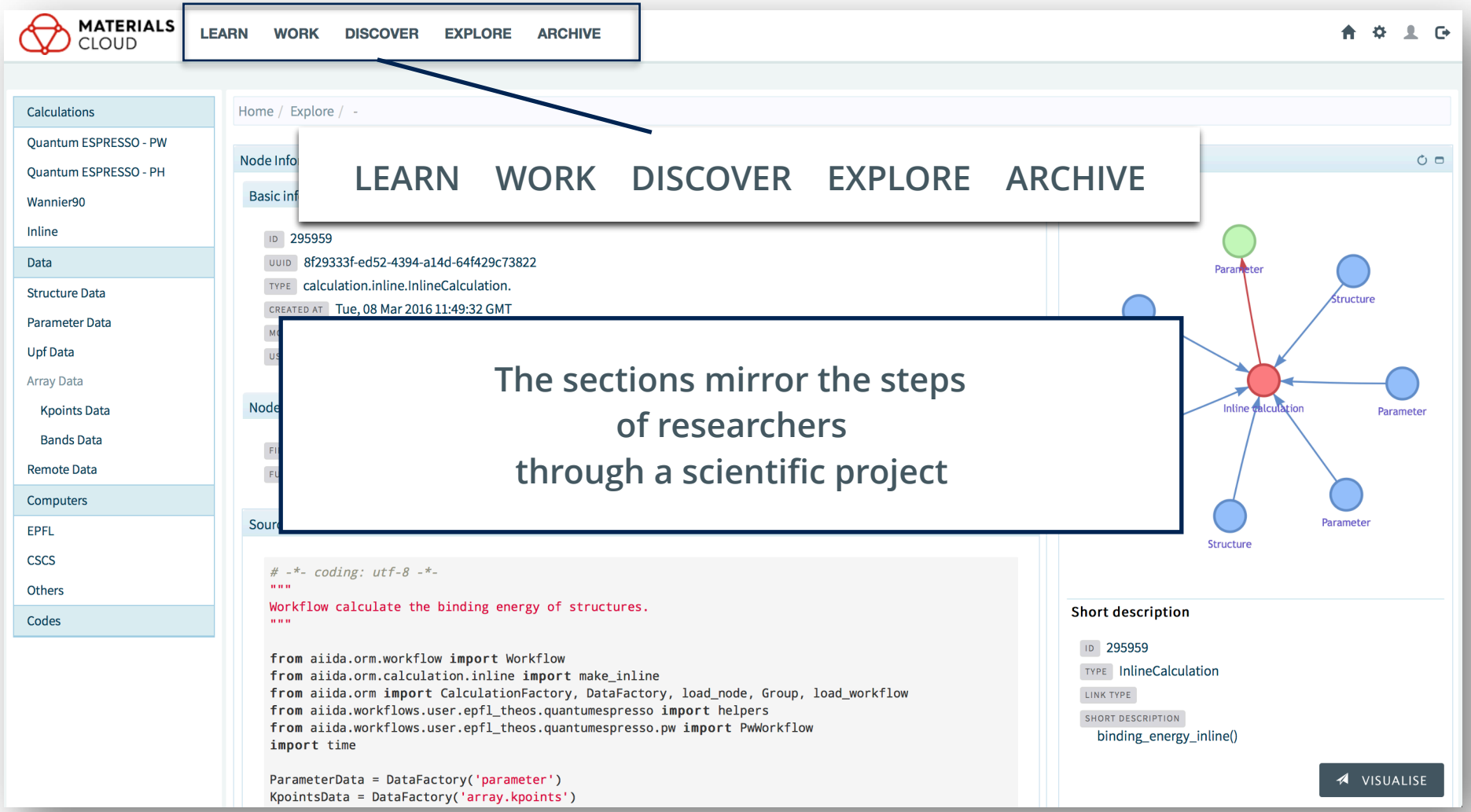
- *Private* AiiDA instances
- UUIDs to uniquely identify nodes

- Data can be shared to other AiiDA repositories **or to online repositories**

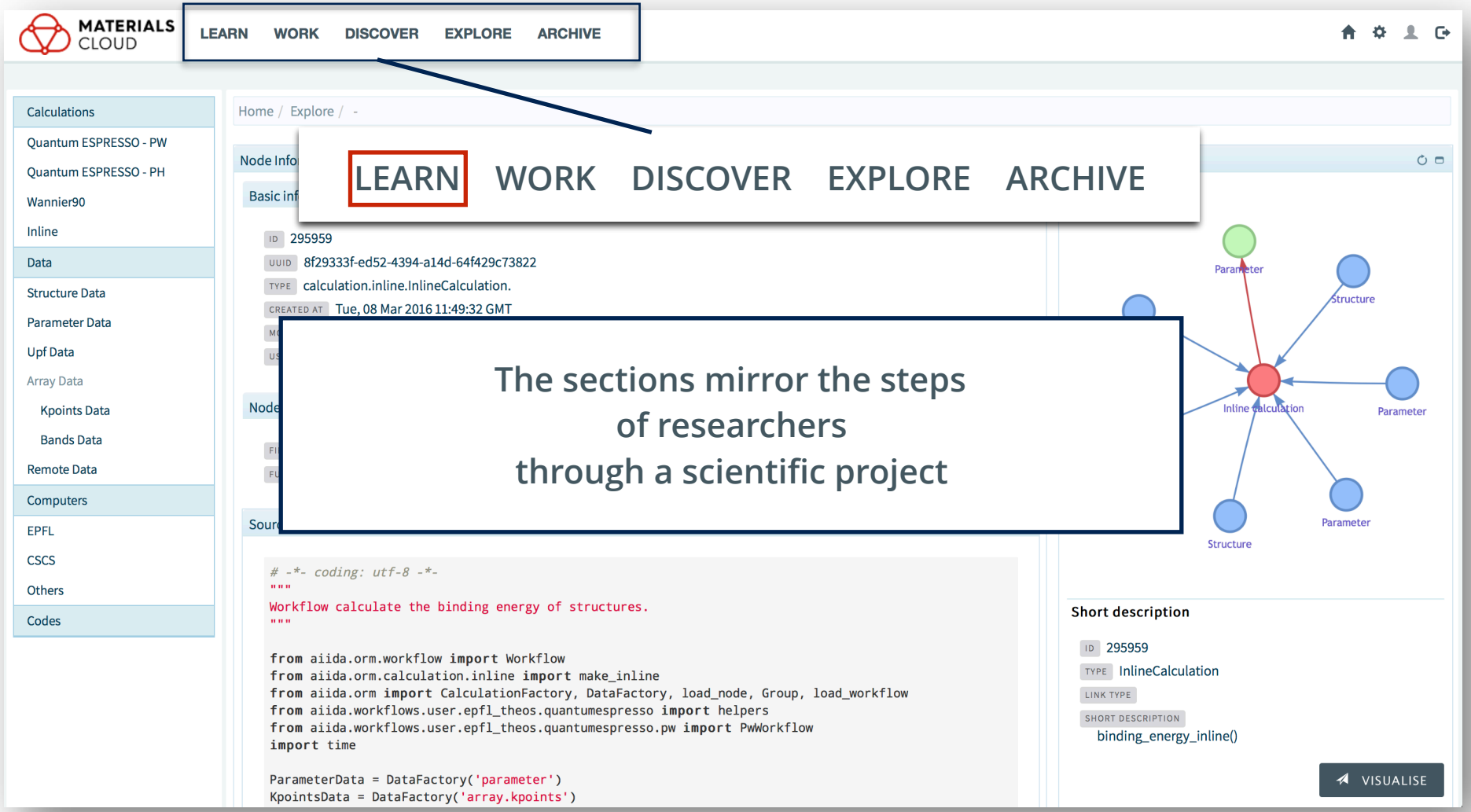


Moving to calculations on the cloud: Materials Cloud (www.materialscloud.org)

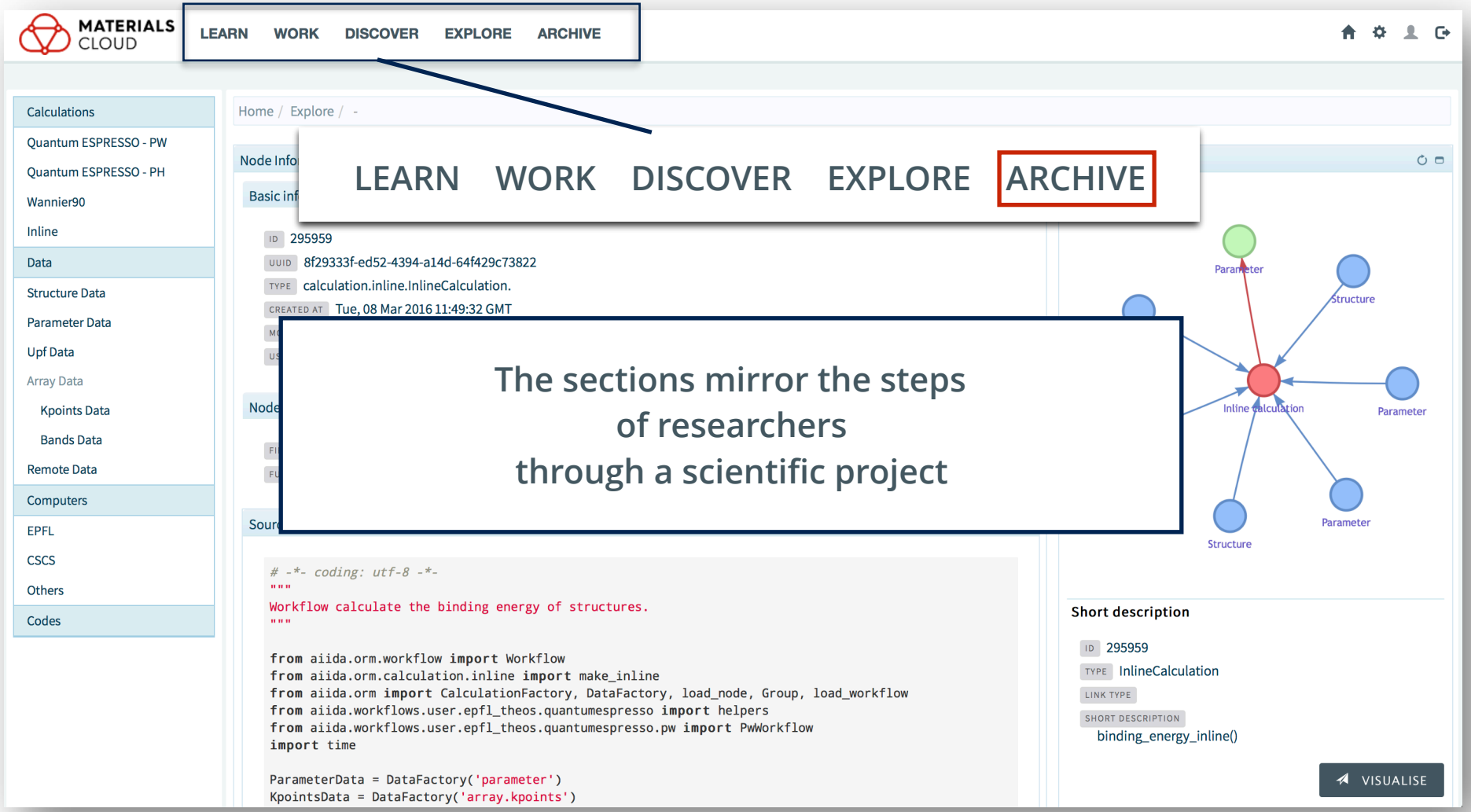
Moving to calculations on the cloud: Materials Cloud (www.materialscloud.org)



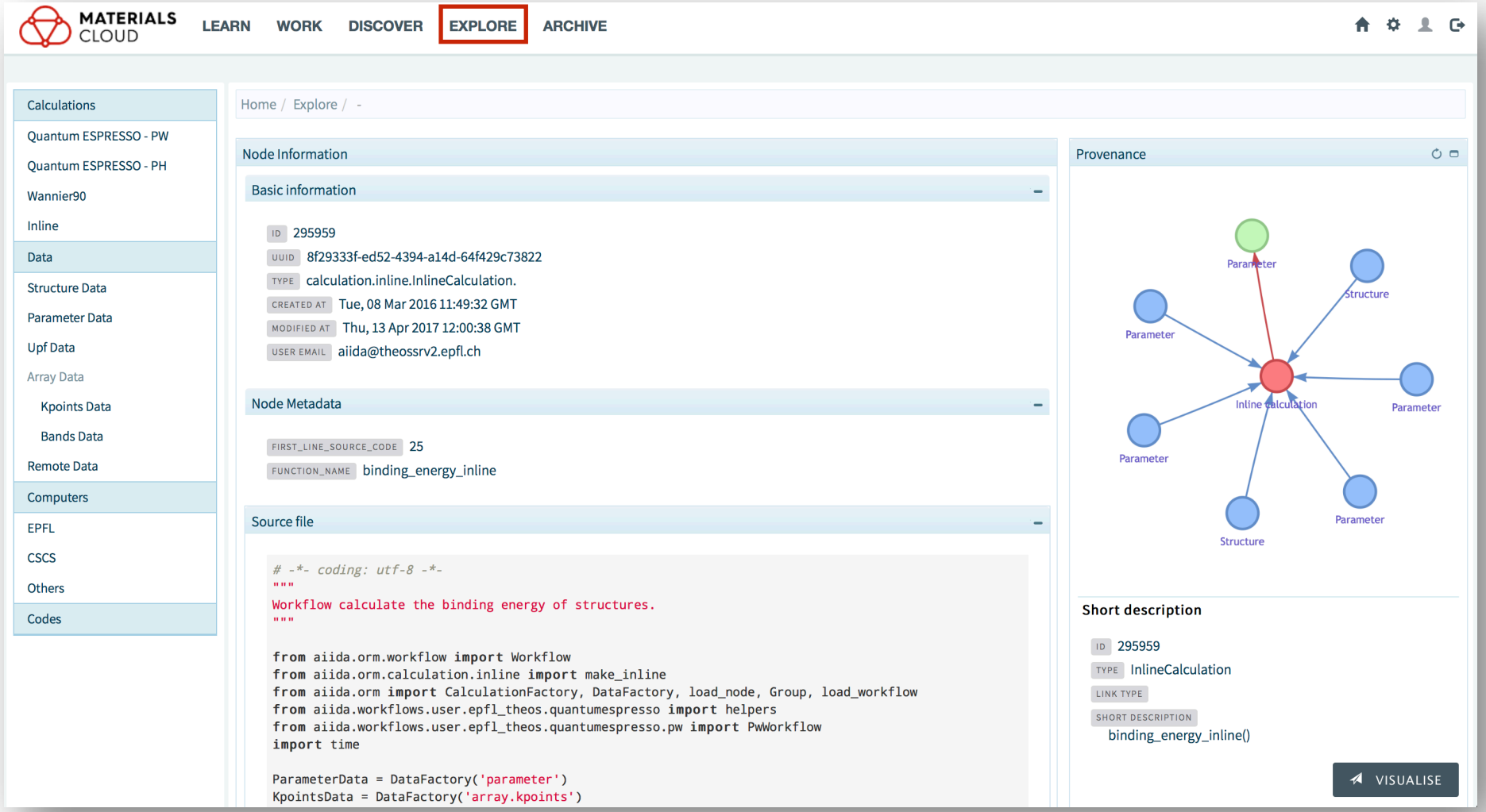
Moving to calculations on the cloud: Materials Cloud (www.materialscloud.org)



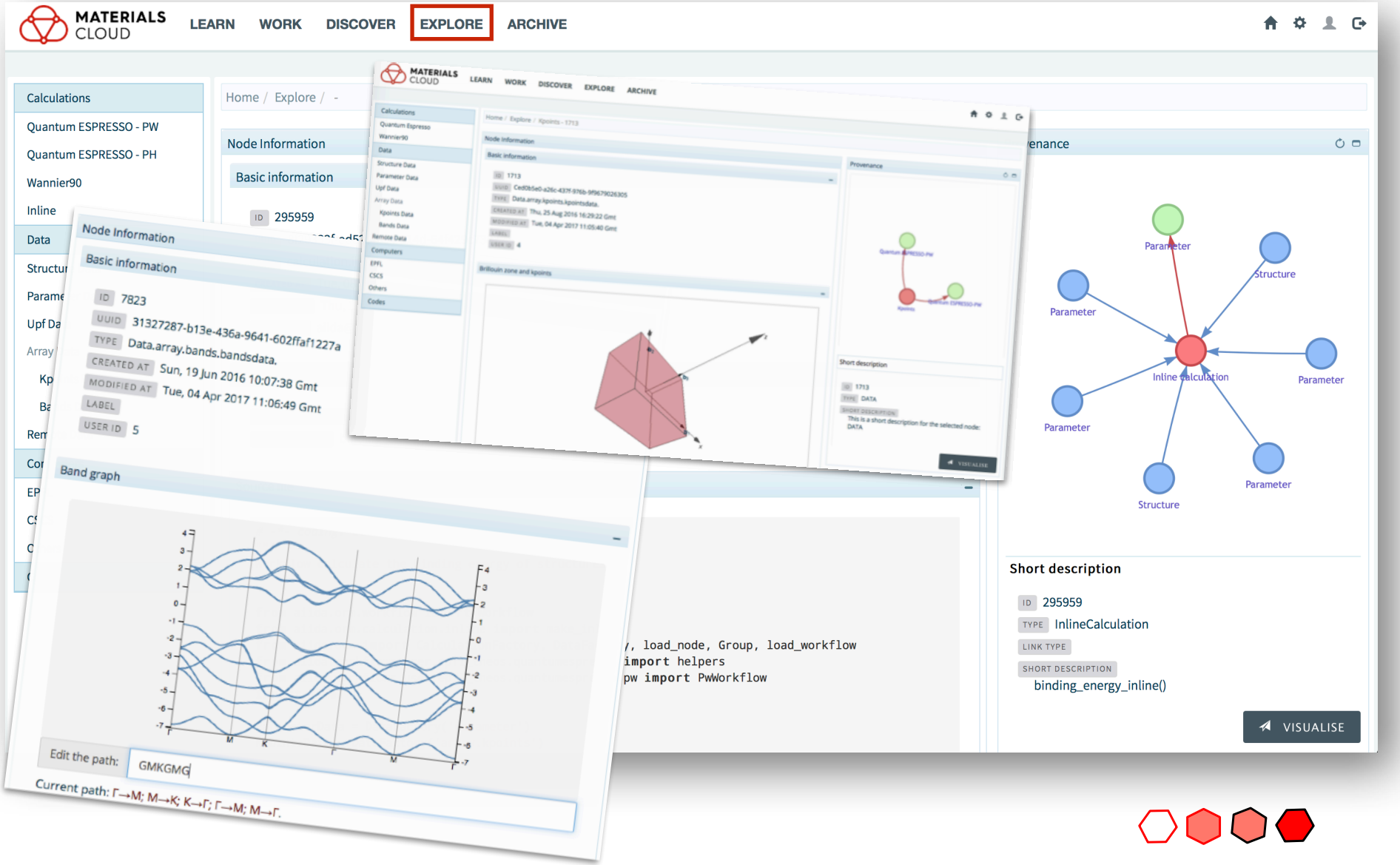
Moving to calculations on the cloud: Materials Cloud (www.materialscloud.org)



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



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



The challenge of data curation (Discover section)

MATERIALS CLOUD | LEARN | WORK | DISCOVER | EXPLORE | ARCHIVE

- Home
- ThermoChemicals
- 2D Structures
- Sketch-map

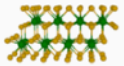
Home / Discover / Menu

Discover home

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

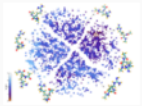
Thermochemical properties of solids

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



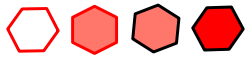
2D structures and layered materials

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)

MATERIALS CLOUD LEARN WORK DISCOVER EXPLORE ARCHIVE

Home
ThermoChemicals
2D Structures

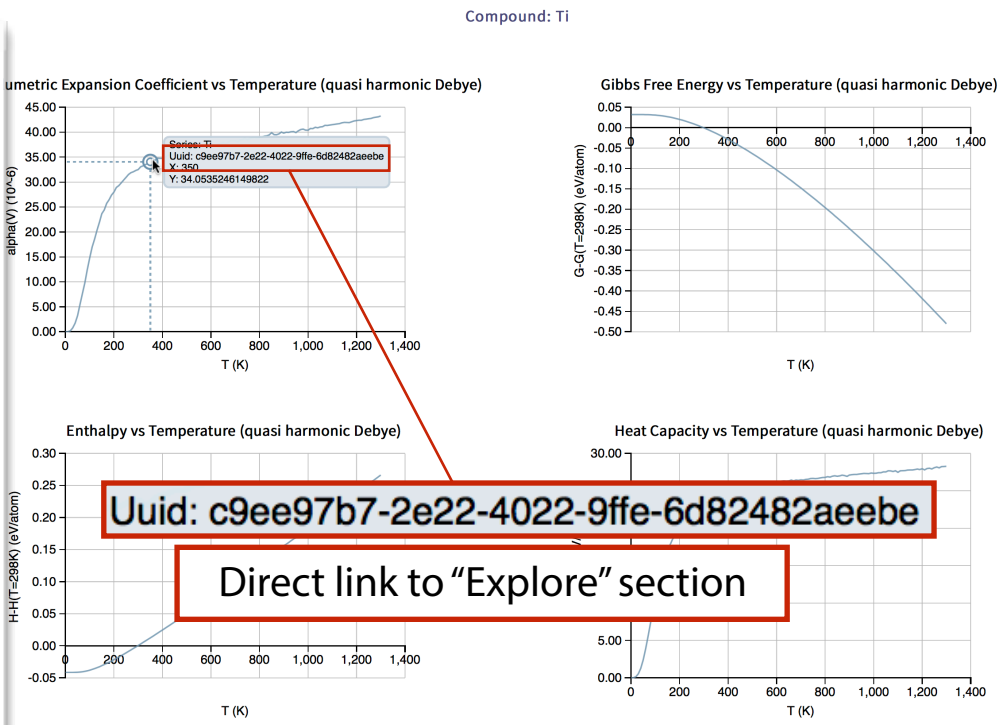
Home / Discover / Menu

Discover home

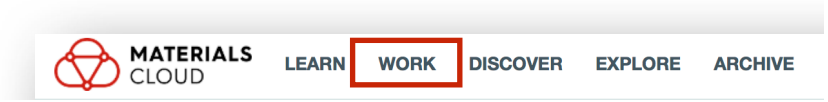
Curated datasets

After raw data generation, but *before* interpretation in a paper

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



Work section: App-store model



"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

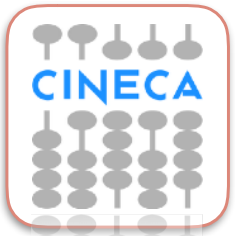

Yambo

CP2K

Fleur

Computing clusters


CSCS (CH)



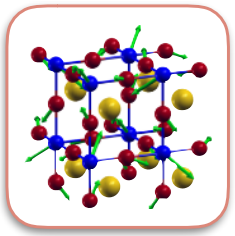
CINECA (IT)

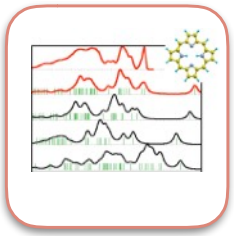
BSC (ES)

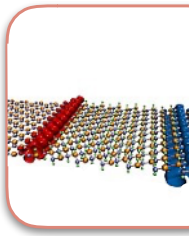

Jülich (DE)

Turn-key workflows

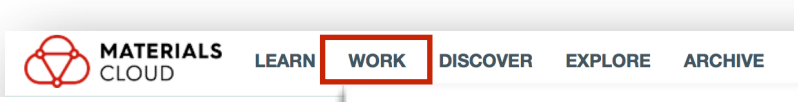

Electronic density


Phonons


Optical properties


Transport

Work section: App-store model



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)

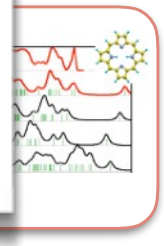
work in collab. with Ole Schütt, EMPA, CH

K

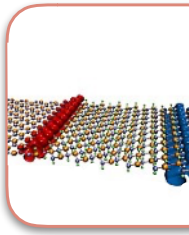
Fleur

(ES)

Jülich (DE)



Optical properties



Transport

Electronic density

Phonons

Acknowledgements

The AiiDA team

Giovanni Pizzi (EPFL)

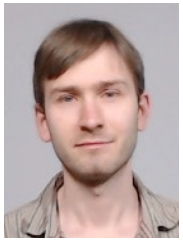
Andrea Cepellotti (EPFL)

Nicolas Mounet (EPFL)

Fernando Gargiulo (EPFL)

Boris Kozinsky (BOSCH)

Nicola Marzari (EPFL)



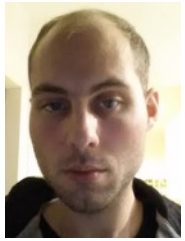
Martin Uhrin (EPFL)

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)

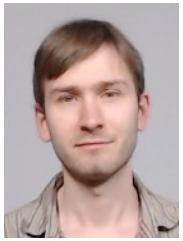
Andrea Cepellotti (EPFL)

Nicolas Mounet (EPFL)

Fernando Gargiulo (EPFL)

Boris Kozinsky (BOSCH)

Nicola Marzari (EPFL)



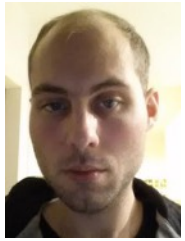
Martin Uhrin (EPFL)

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

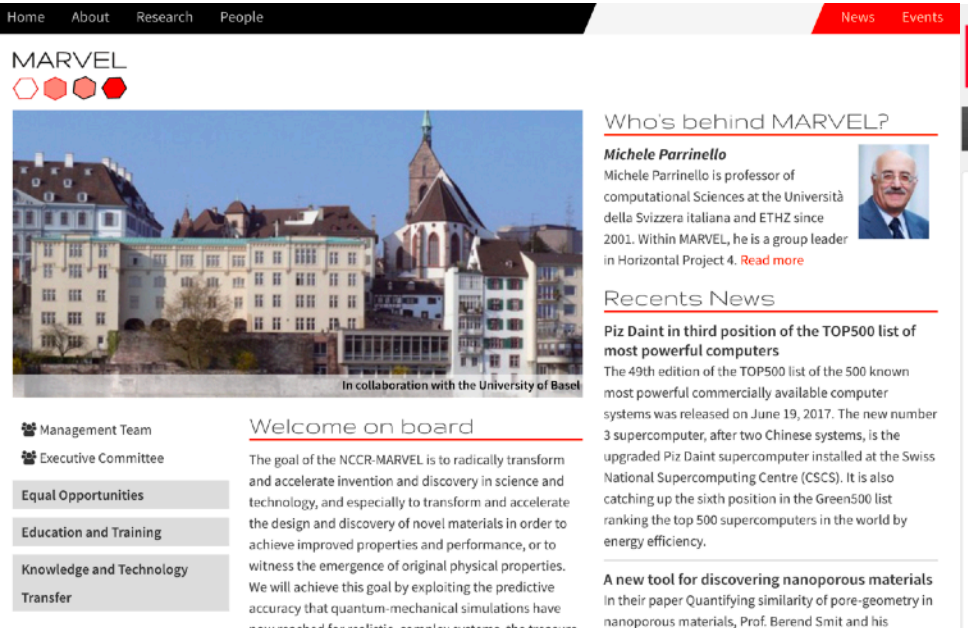
41 PIs from 12 Institutions
(including the Swiss supercomputing center CSCS)
http://marvel-nccr.ch

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



Conclusions

A computational science platform adopting the ADES model:

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