RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

FRANCE 2030
PROGRAMME DE RECHERCHE
MATÉRIAUX ÉMERGENTS

anr

cnrs

Accelerated Design

Databases & AI

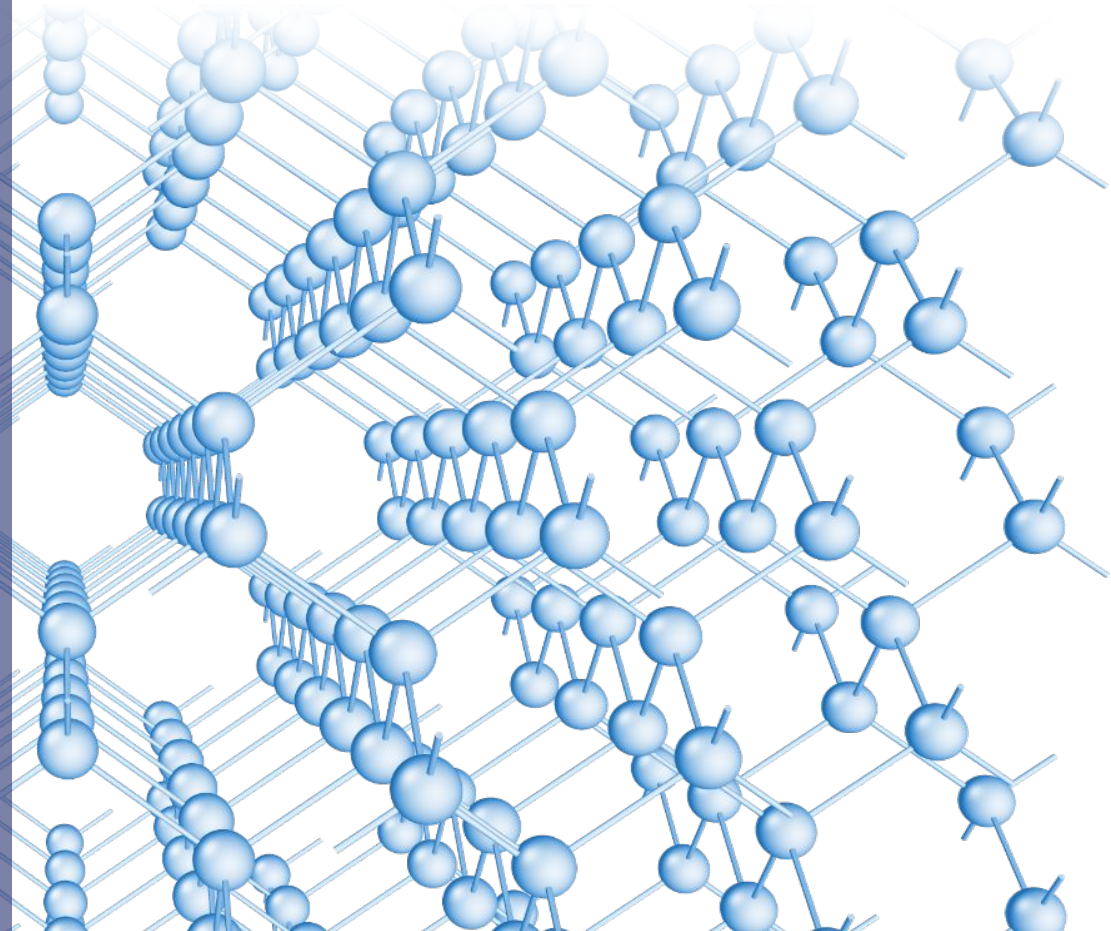
Codes & Workflows

DIAMOND : ANR-22-PEXD-0015

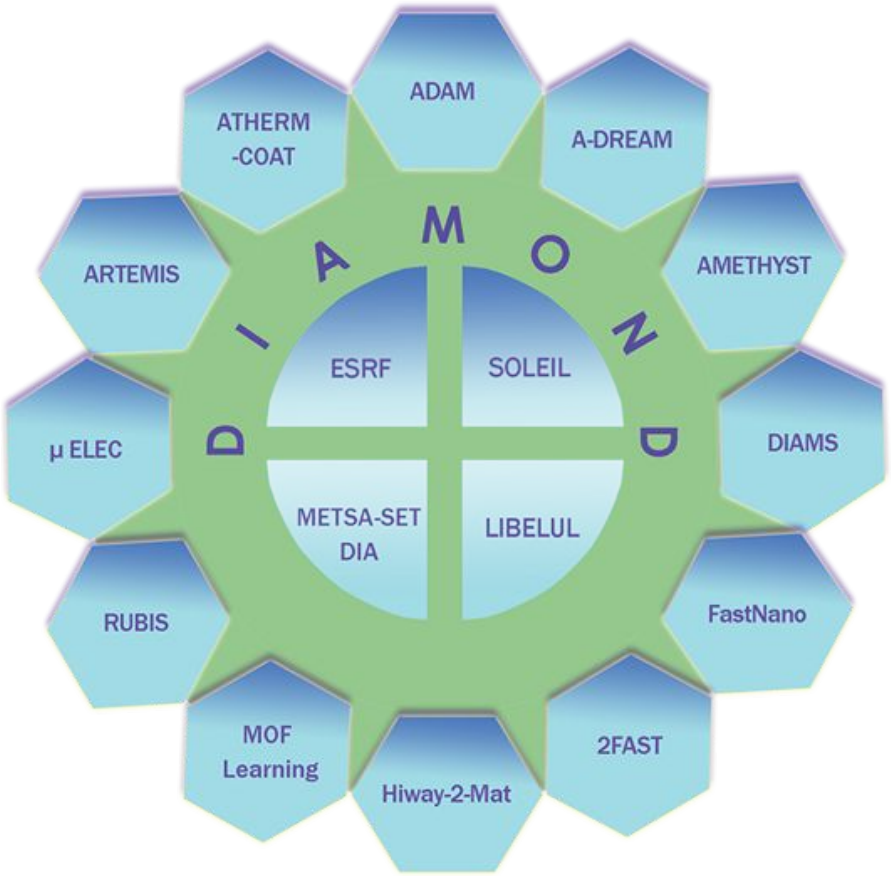
Journée utilisateurs Gricad 2025

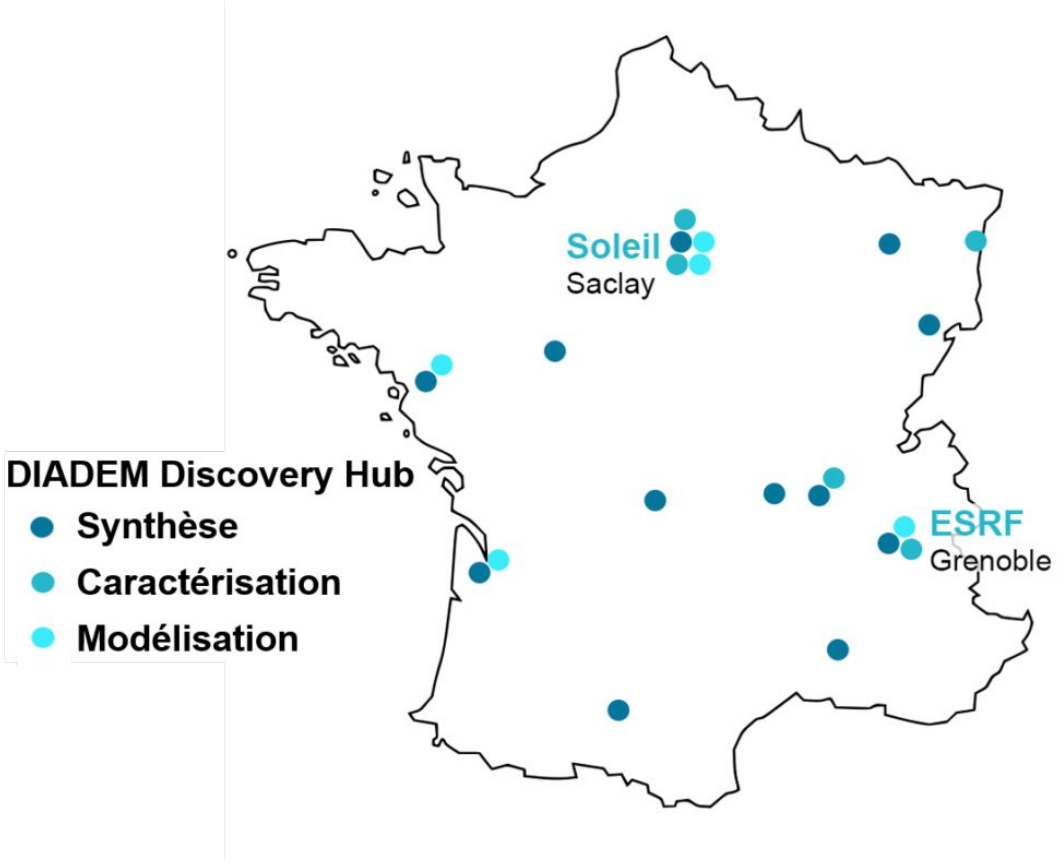
Diamond platform

For accelerated materials design

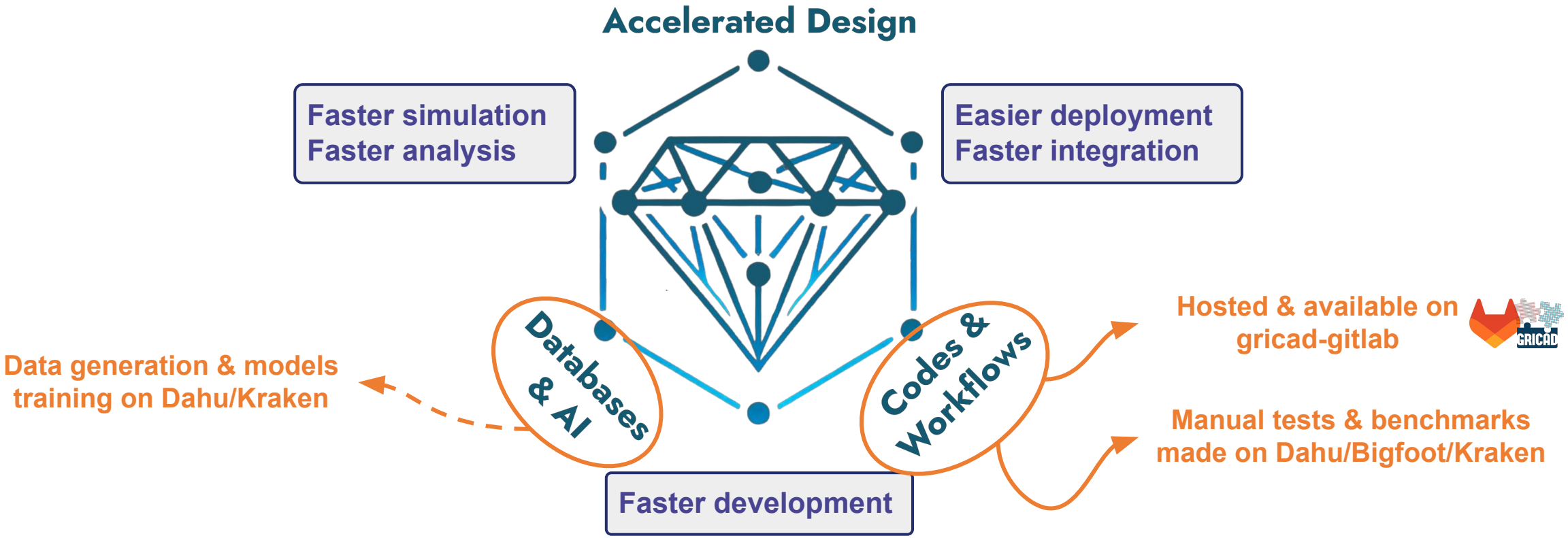


DIADEM Projects & Platforms





Diamond, the numerical platform



Containerization, a software need

¯_(ツ)_/¯

It works on my machine

Problem

> 1h of installation & debugging everytime someone want to try it

Micro-services / Personal use

High Performance Computing

Containerization issues

build : Aug, 8th

build : Nov, 6th

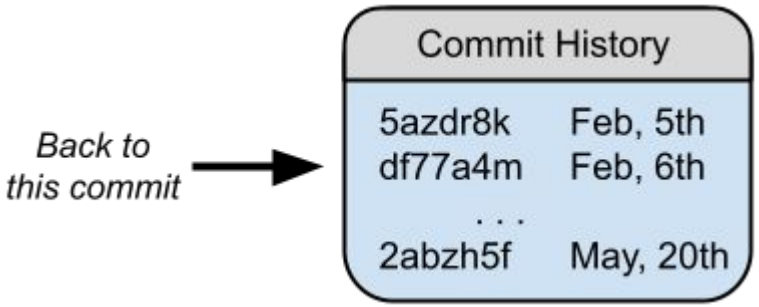
Package differences can cause **diverging results** and/or **errors**!

Reproducibility, a scientific need

The ability to reproduce the environment of software that accompanies research work should not be considered a mundanity or an exercise that's "overkill". The ability to rerun, inspect, and modify software are the natural extension of the scientific method. Without a companion reproducible software environment, research papers *are merely the advertisement of scholarship*, to paraphrase Jon Claerbout.

- **Ludovic Courtès,** *Maintainer of Guix*

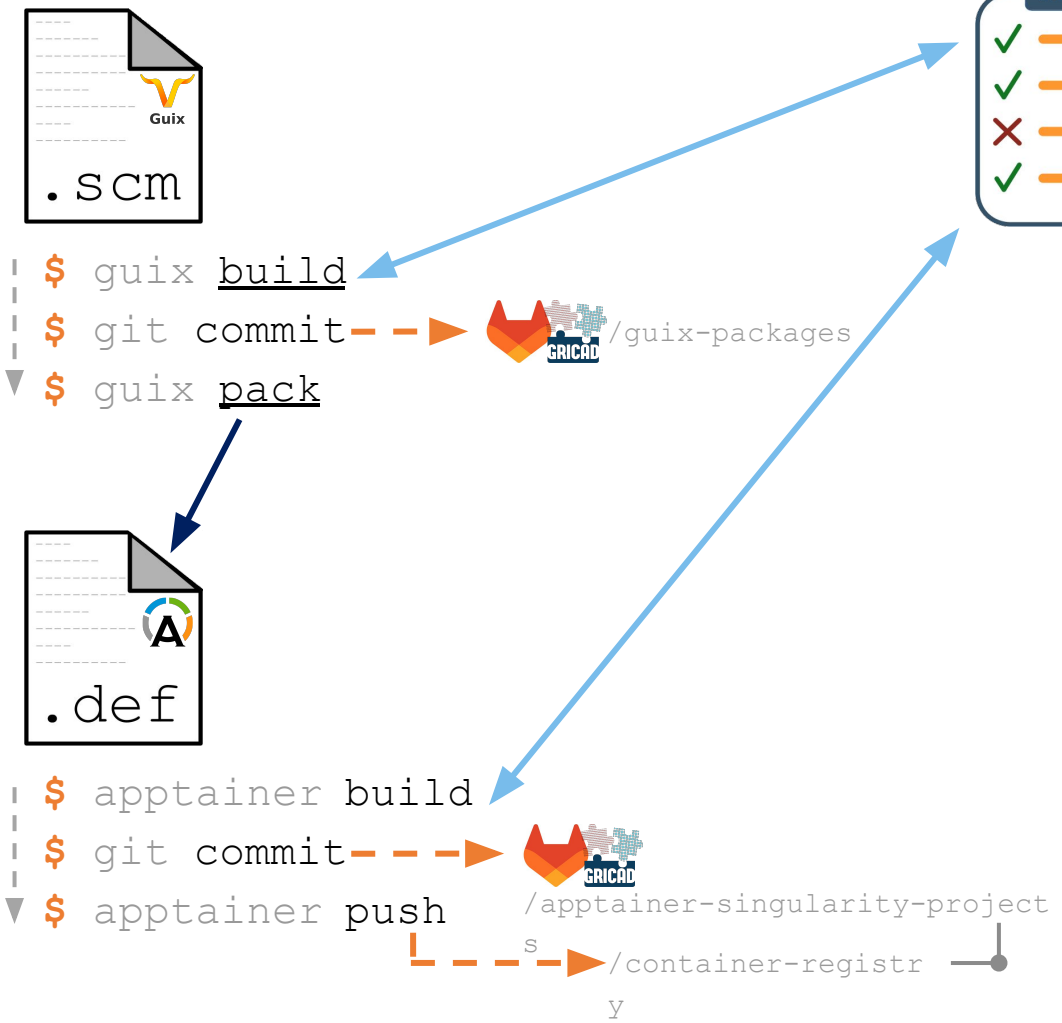
Guix

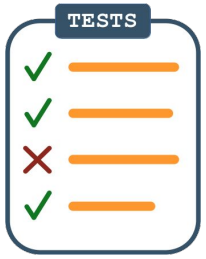



```
$ guix install python-numpy
$ guix time-machine --commit=df77a4m -- install python-numpy
```

Every dependency is frozen as it was at a given commit time

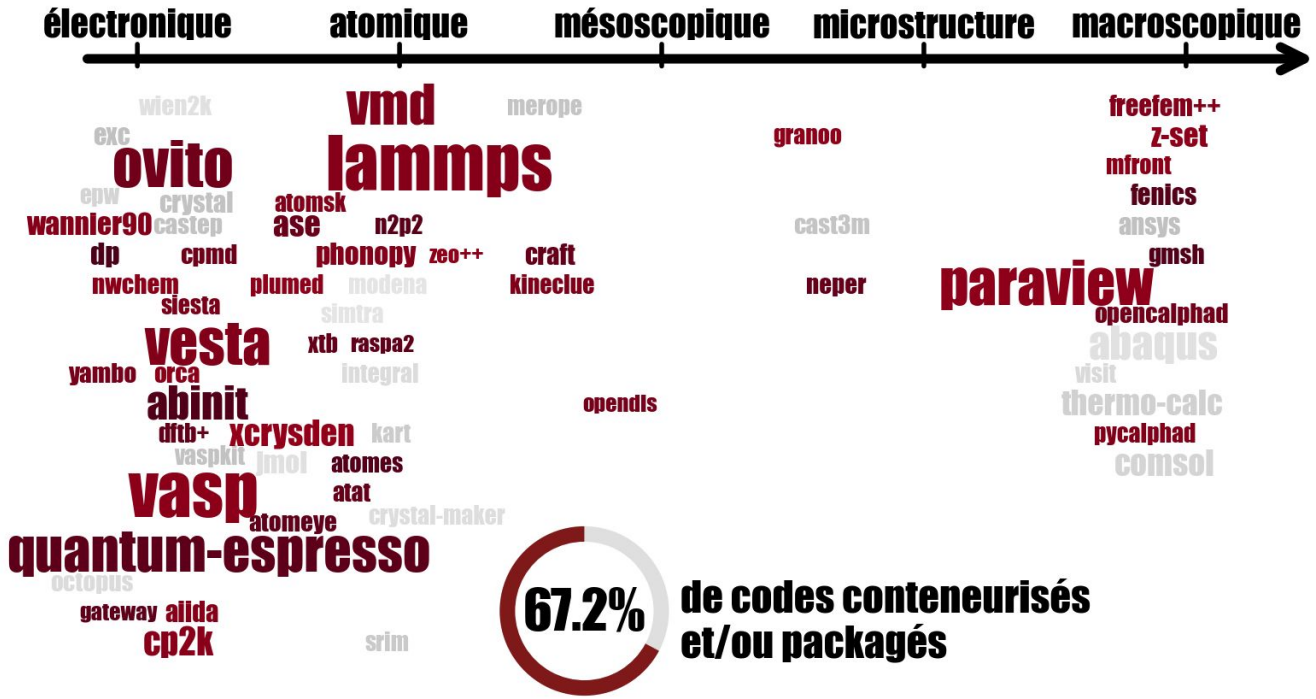
Containerization-packaging process





Tests can be varied (official QA, public tutorials, …)

→ **manual** checking can be needed !



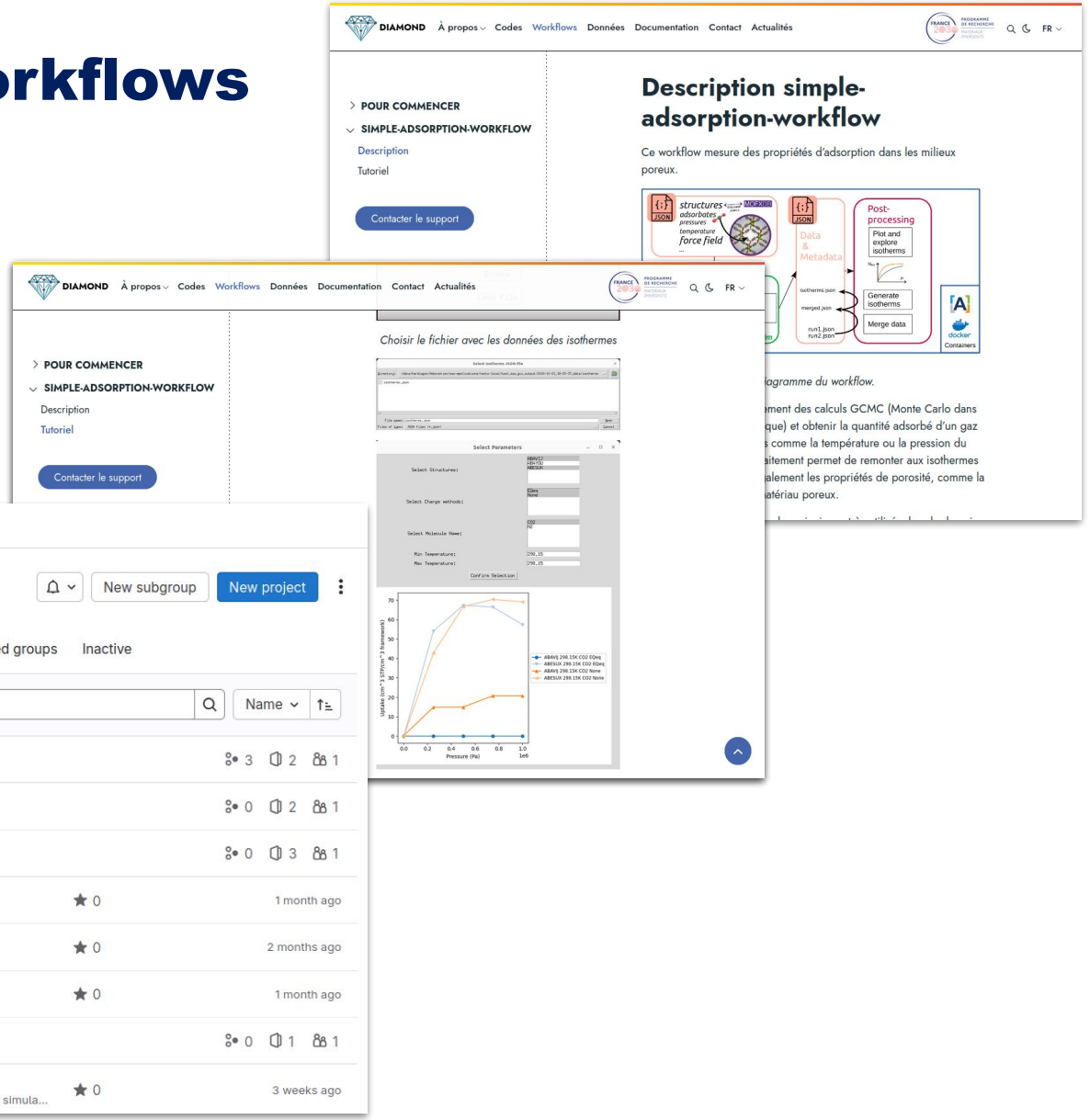
de codes conteneurisés et/ou packagés

Access DIAMOND codes & workflows

- A **dedicated space in the website** was set to host the workflows and codes documentations.
- The **first demonstrator workflow** of MOFLearning (a Python workflow for gas absorption simulations with RASPA and Zeo++) is **already public**, with tutorials and benchmark accessible via our website.

- As soon as the workflows give rise to **scientific contributions** and **be published**, they will be made available via our **GRICAD Gitlab**, at the same address as all our other tools and plugins.

- Open source code are available **directly** after we manage to package them



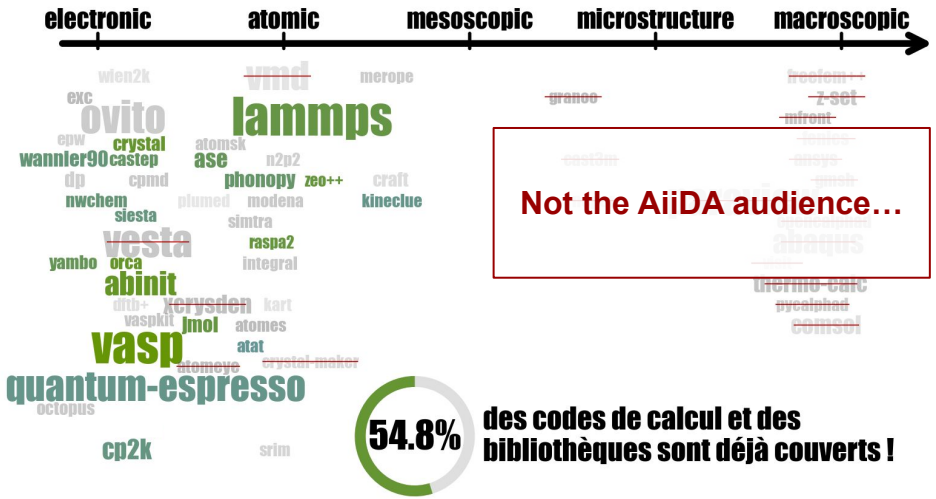
Why use AiiDA ?

Covalent Parsl mlflow AiiDA pyiron

and many more …

Choice motivated by :

- Proximity to AiiDA developers at SIMaP
- The list of existing plugins
- The list of existing schedulers



In practice, AiiDA relies in plugins

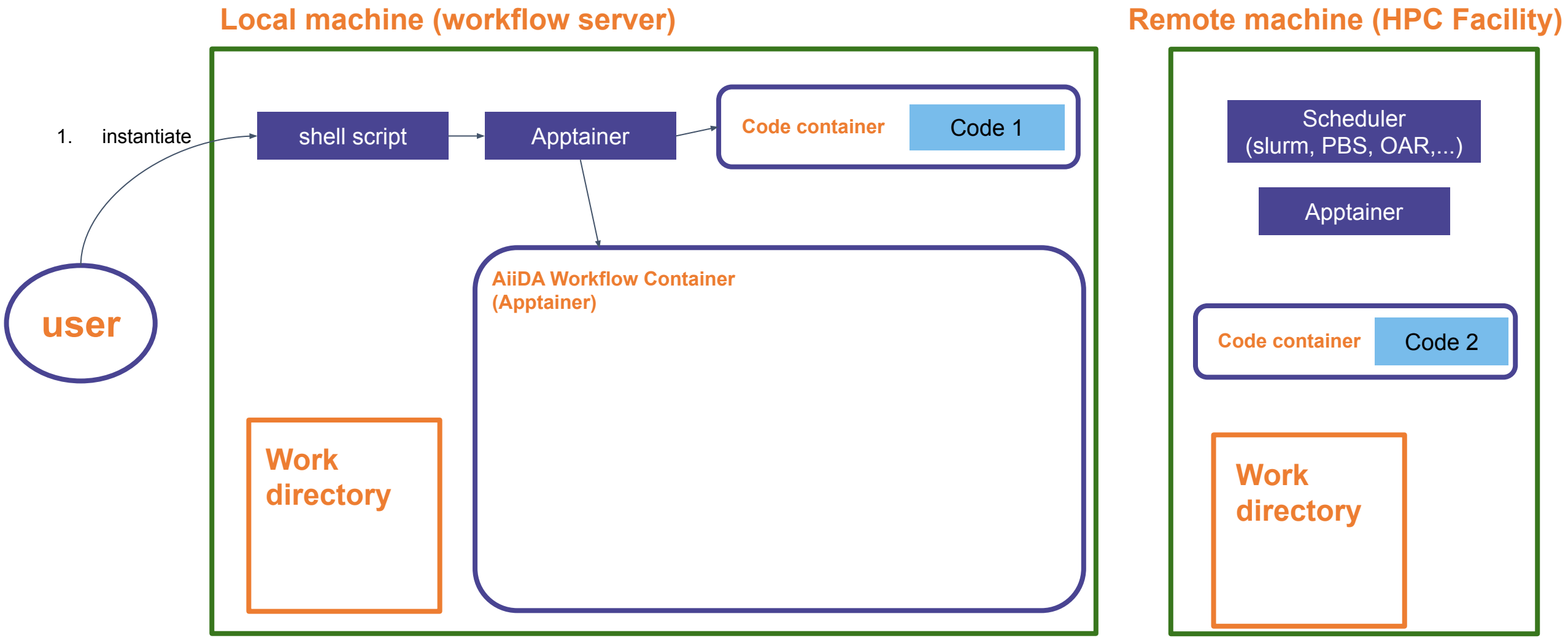
- ***for codes**, to to communicate with simulation software*
- ***for schedulers**, to deploy calculations in HPC facilities*
- ***for communication protocols**, to talk between different machines performing a task*

Here is where we did most of the technical work! We did: aiida-oar, fifo-transport, aiida-n2p2, aiida-numodis… Also forks of aiida-lammps and aiida-vasp with customizations to better serve our workflows.

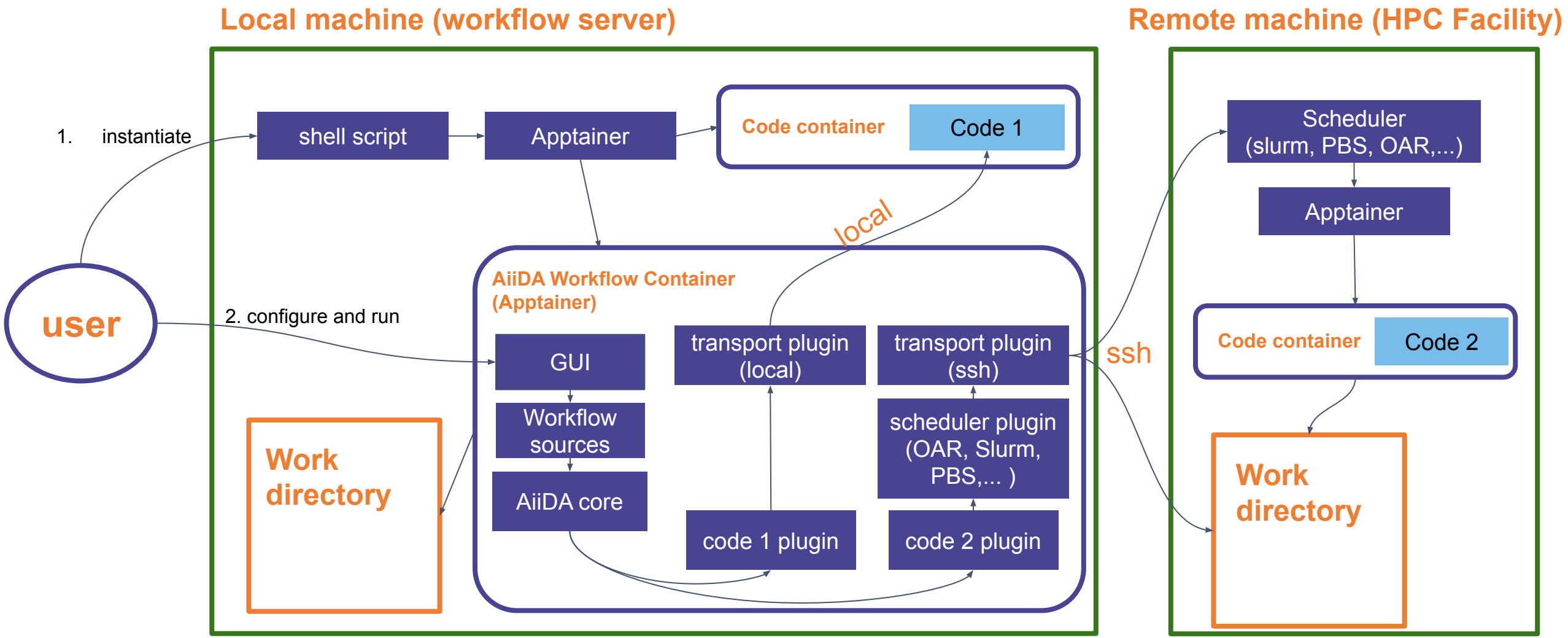
GRICAD SIMaP cnrs cea

AiiDA workflow

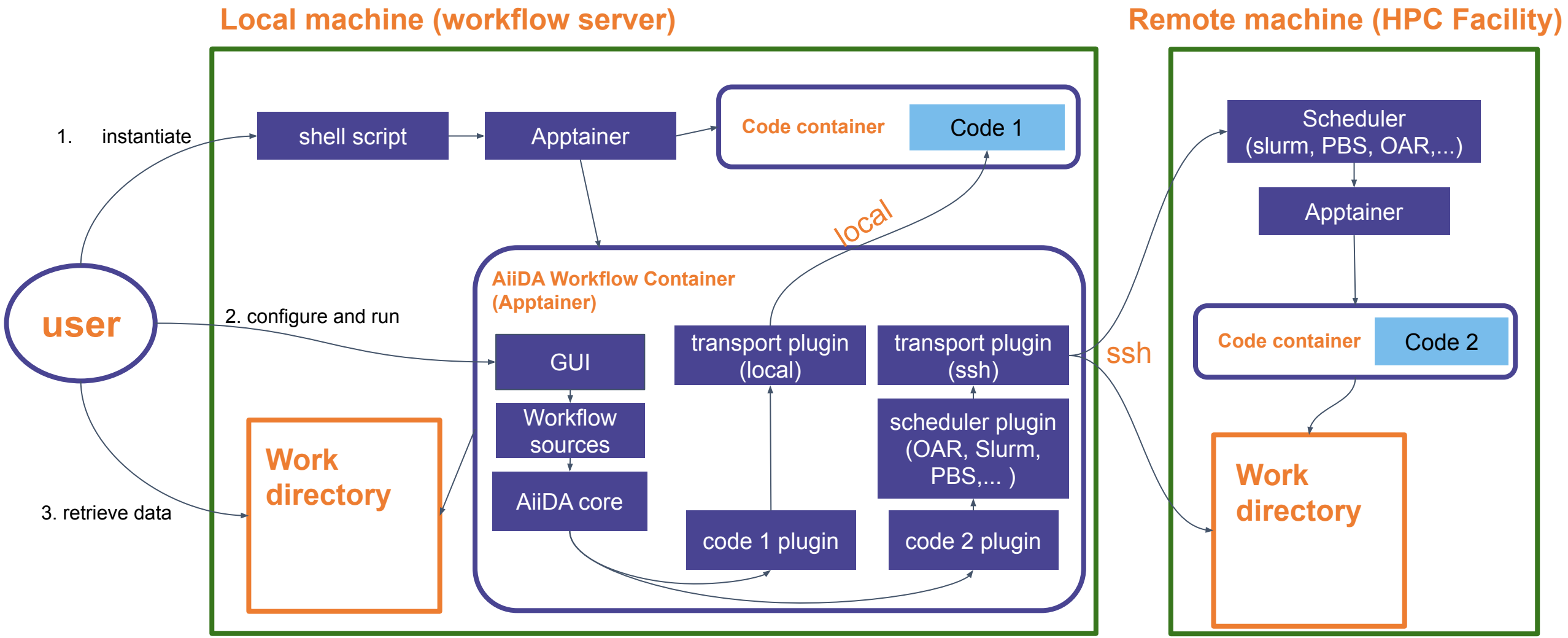
AiiDA workflow



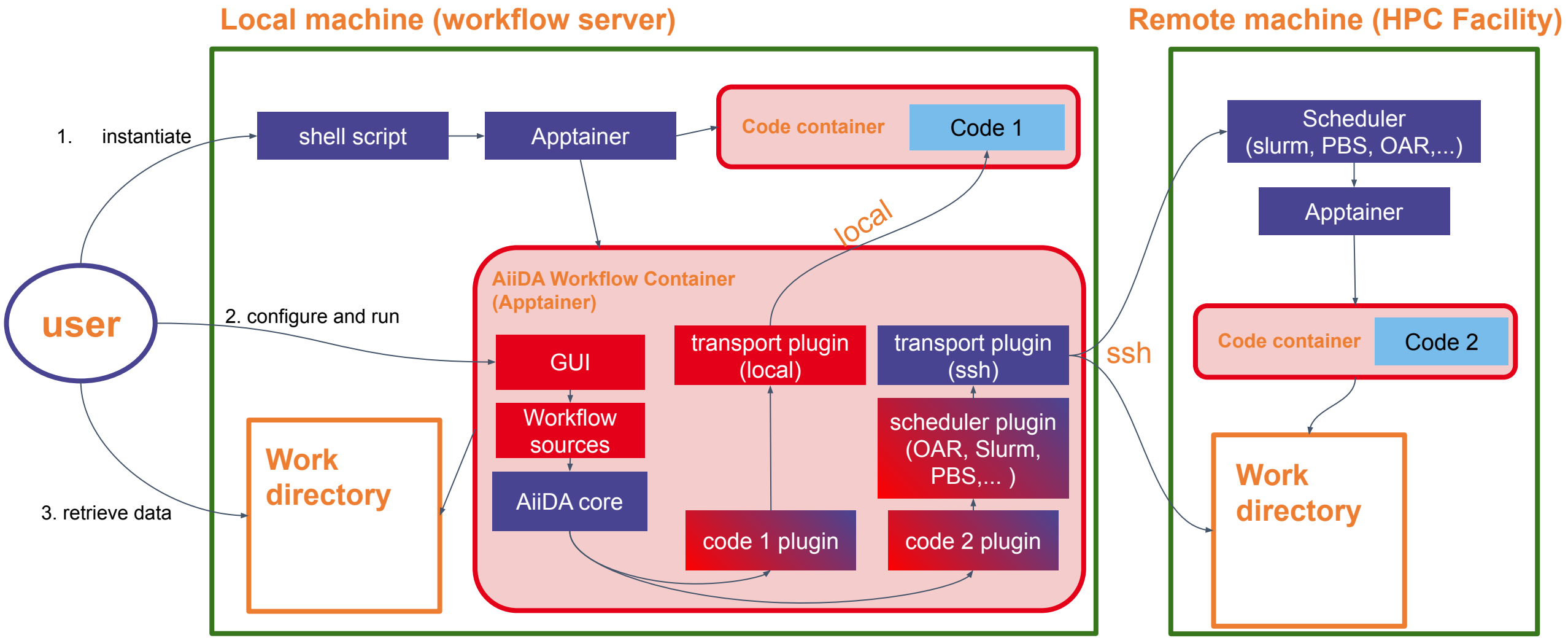
AiiDA workflow



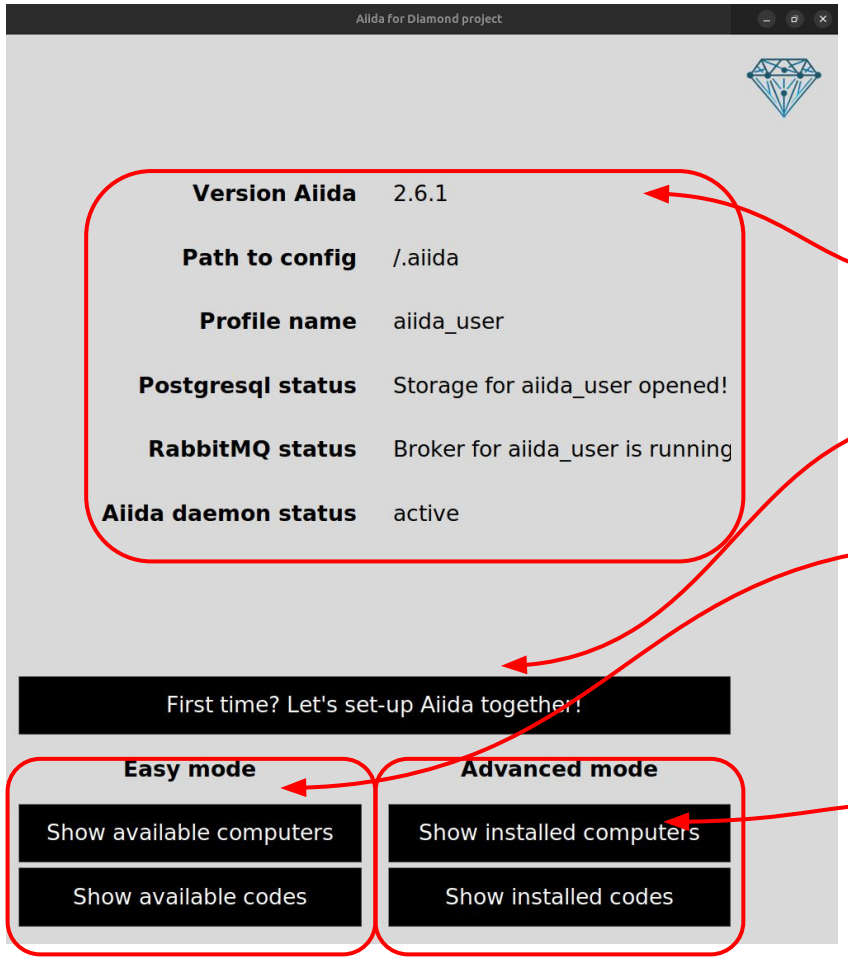
AiiDA workflow



AiiDA workflow



Development of GUI for AiiDA Workflow Manager



The main objective is to make using AiiDA **as easy as possible**!

Multiple views for all users :

- **Information Panel**, displaying current AiiDA setup
- A **Beginner Mode**, building fast setup, ready in a click!
- An easy click-and-play menu to install codes or computers from **pre-made configuration files:** *[VASP, LAMMPS, QE,...] and [DAHU, JEAN-ZAY,...]*
- An **Advanced Mode** to those who want to edit details of even delete codes or computers already installed

Our Workflows:

	Scientific specialists:	Status of development:
• **NUMODIS+LAMMPS workflow** → Atomistically informed dislocation dynamics simulations	Laurent Dupuy	Stable, Containerized, Private
• **MOFLearning workflows** → Evaluate gas- absorption properties of MOFs, screening large databases	Arthur Hardiagon François-Xavier Coudert	Stable, Containerized, Private, Production
• **Quantum Espresso workflow** → Screening of materials for electronic structure related properties	Ashna Jose Mattia Ragni Roberta Poloni Martin Uhrin	WIP
• **VASP workflow** → Procedural AIMD trajectories aiming the training MLIPs for metals	Jonathan Chapigna Noël Jakse	WIP, Containerized, Private, Production
• **MLIP workflow - n2p2** *[WP3]* → Automate the training and validation of MLIPs with the n2p2 package	Arthur France-Lanord Marco Saitta	WIP, Private

Workflows VASP

A Workflow for liquid alloys MD via VASP

- **Implementation and analysis:** J P A Mendonca, N Jakse
- **Technical support / containerization:** B Arrondeau, D Bissuel, P-A Bouttier

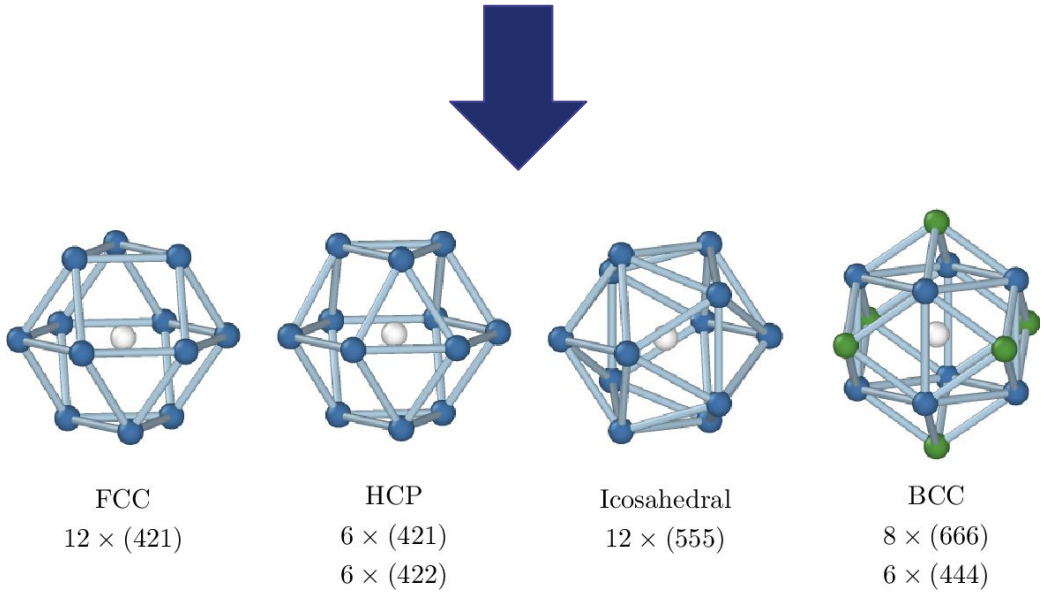
- Initially proposed by Noel and P-A as a **test case** for the platform, to evaluate the automation limits and guide our development efforts (martyr WF)
- Some more tangible interests in this workflow:
  - There is **MD data produced by Noels group** that still needs time expensive (human time) exploitation. E.g., doing structural evolution analysis in MDs with about 20k atoms and more the 1k non correlated frames.
  - The automated workflow can be extremely **useful for WP3**, creating datasets for future MLIPs.
  - There are at least two people in Noels group that can be very interested and active **beta testers**
- We are starting with AiiDA, but one of the goals is also to **reproduce it with other WF managers**, to test their limits and capabilities

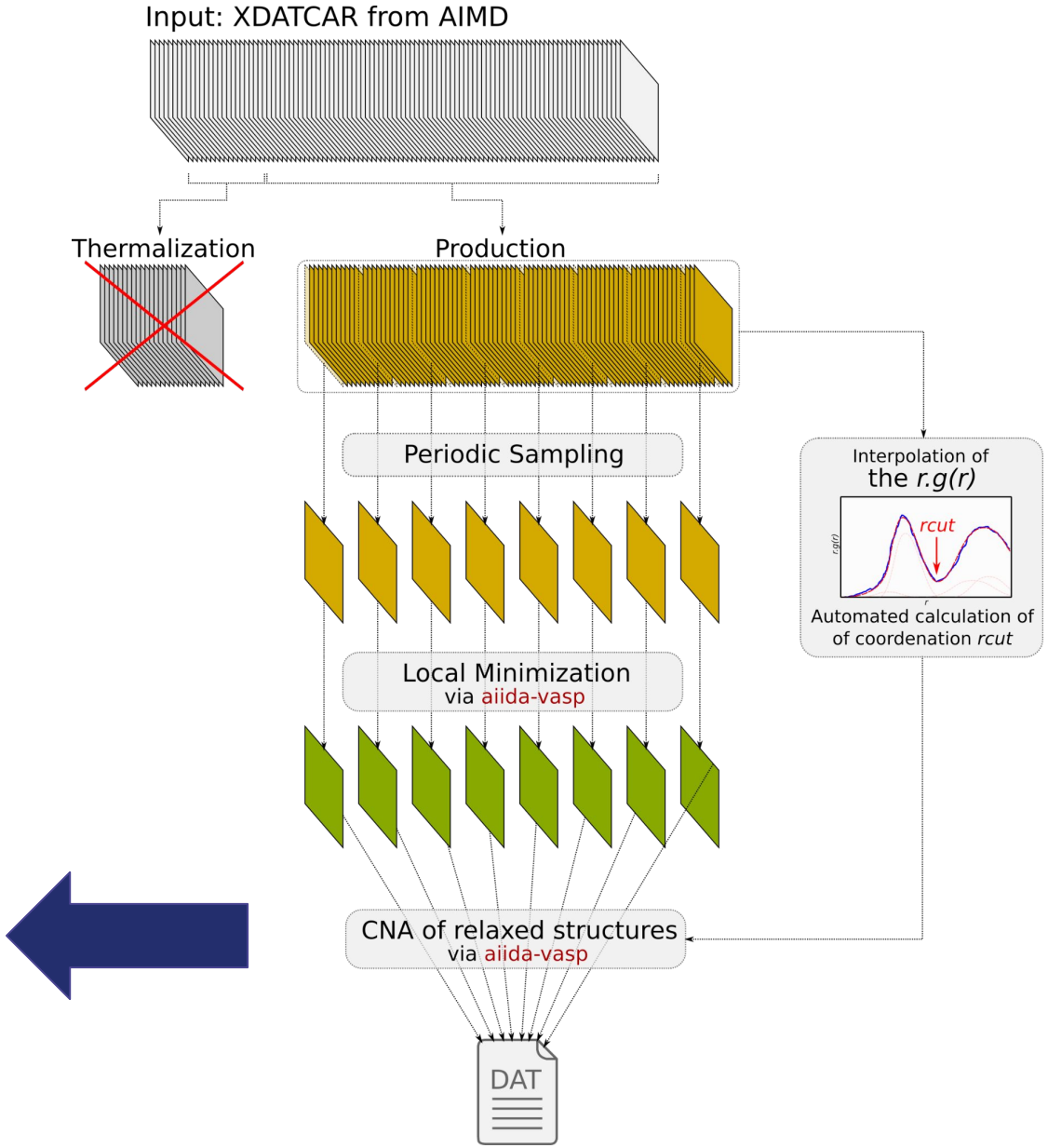
Workflows VASP - Module 1

Workflows organization:

WF1 - MD structural analysis

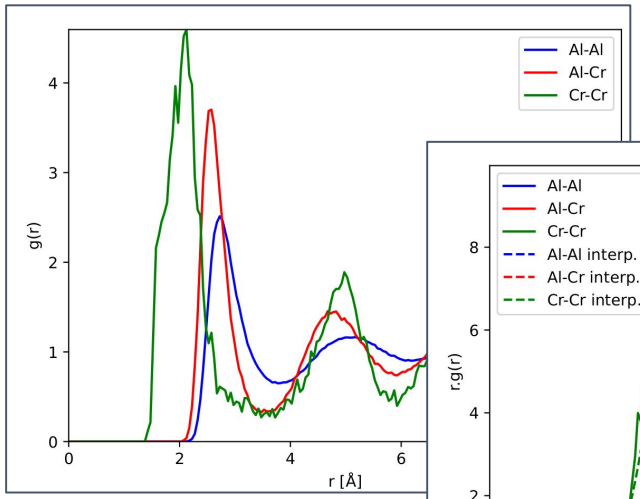
- Input: MD trajectory from VASP
- Process: Sampling > Structure Relaxation > Structural Analysis
- Output: Inherent Structures and CNA analysis (per atom / global)



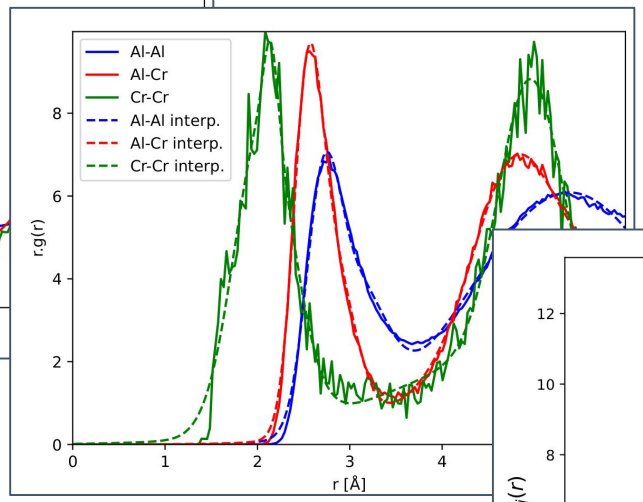


Workflows VASP - Module 1

g(r) from MD

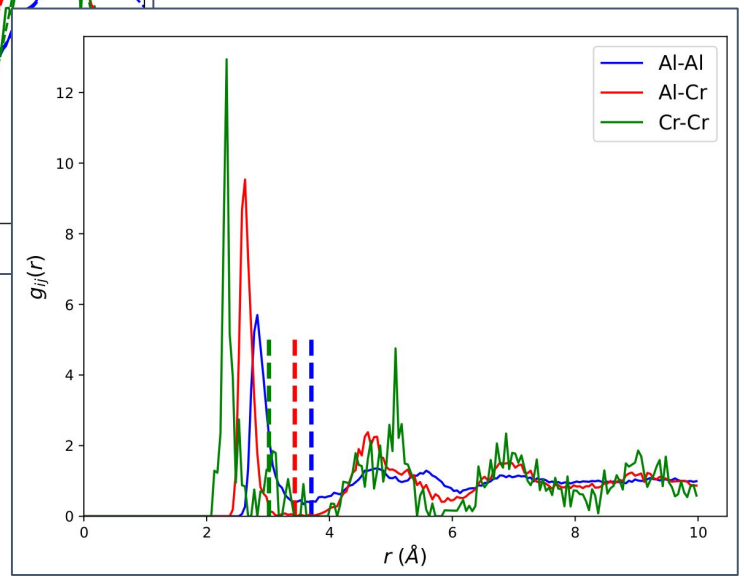


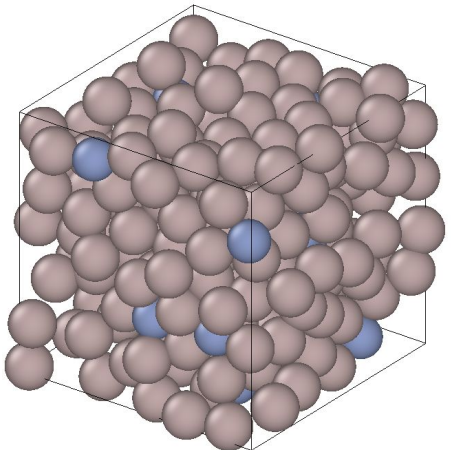
r.g(r) from MD



Each structure to optimise
= 64 cores for 24h

g(r) inherent structures





$Al_{238}Cr_{18}$

CNA Analysis

```
-----------------FILE  POSCAR_8000 ----------------------

POSCAR_8000

1 1 13 [3 0 0]:1 [4 2 1]:1 [4 2 2]:3 [4 3 3]:3 [5 4 4]:3 [5 5 5]:2
2 1 10 [2 0 0]:3 [3 1 1]:2 [3 2 2]:1 [4 2 2]:2 [4 3 3]:1 [5 4 4]:1
3 1 12 [3 0 0]:2 [3 1 1]:1 [3 2 2]:1 [4 2 2]:3 [4 3 3]:3 [5 4 4]:1 [5 5 5]:1
...
254 2 10 [4 3 3]:3 [4 4 4]:1 [5 4 4]:1 [5 5 5]:5
255 2 12 [4 3 3]:2 [5 4 4]:2 [5 5 5]:8
256 2 10 [3 1 1]:2 [4 3 3]:6 [5 5 5]:2

Total number of pairs for atom type 1: 2632
Total number of pairs for atom type 2: 196
Total number of pairs for atom type 3: 0
Total number of pairs for atom type 4: 0

Number of atoms type 1: 238
Number of atoms type 2: 18
Number of atoms type 3: 0
Number of atoms type 4: 0

Coordination number type 1: 11.058823529411764
Coordination number type 2: 10.88888888888889
Coordination number type 3: 0.0
Coordination number type 4: 0.0
Coordination number total : 11.046875

 CNA particle type 1
[1 0 0]:18 0.006839
[2 0 0]:206 0.078267
[2 1 1]:34 0.012918
[3 0 0]:228 0.086626
[3 1 1]:484 0.183891
[3 2 2]:144 0.054711
[4 1 1]:42 0.015957
[4 2 1]:132 0.050152
[4 2 2]:418 0.158815
[4 3 3]:485 0.184271
[4 4 4]:4 0.001520
[5 3 2]:4 0.001520
[5 3 3]:4 0.001520
[5 4 4]:238 0.090426
[5 5 5]:180 0.068389
[6 6 6]:11 0.004179

 CNA particle type 2
[3 0 0]:2 0.010204
[3 1 1]:6 0.030612
[4 2 1]:6 0.030612
[4 2 2]:14 0.071429
[4 3 3]:57 0.290816
[4 4 4]:4 0.020408
[5 4 4]:22 0.112245
[5 5 5]:84 0.428571
[6 6 6]:1 0.005102

 CNA particle type 3

 CNA particle type 4
```

Chemical bound close to crystallization:

6.8% Al

42.9% Cr

Workflows VASP - Module 2

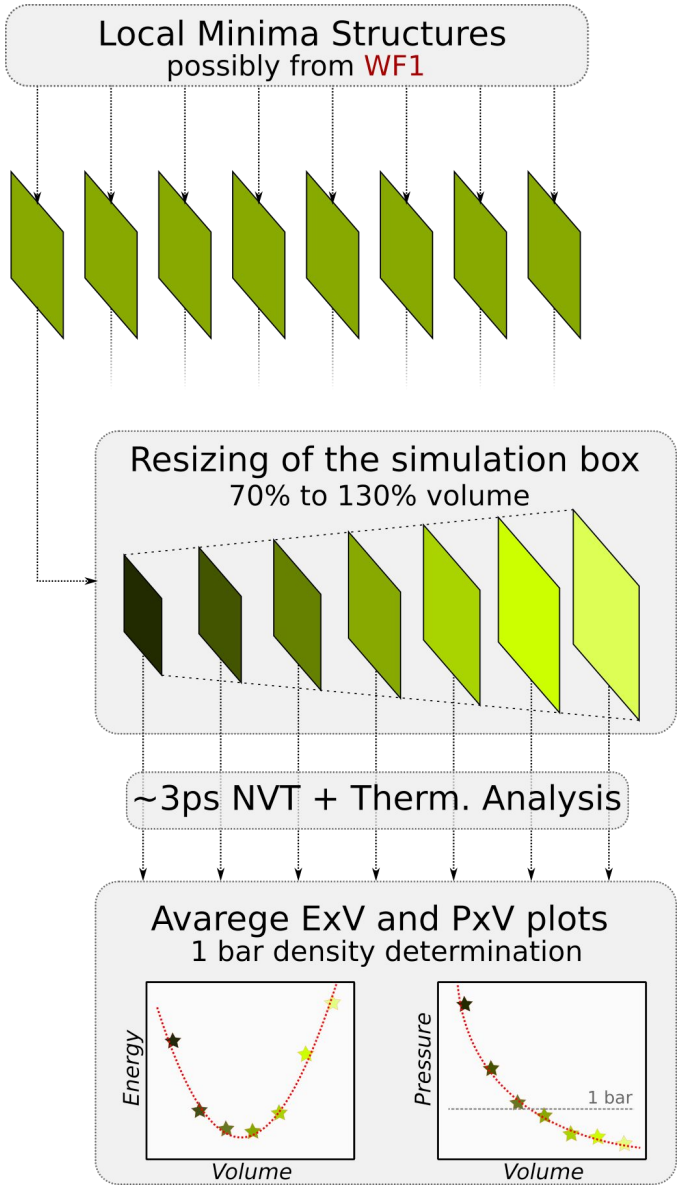
Workflows organization:

WF1 - MD structural analysis

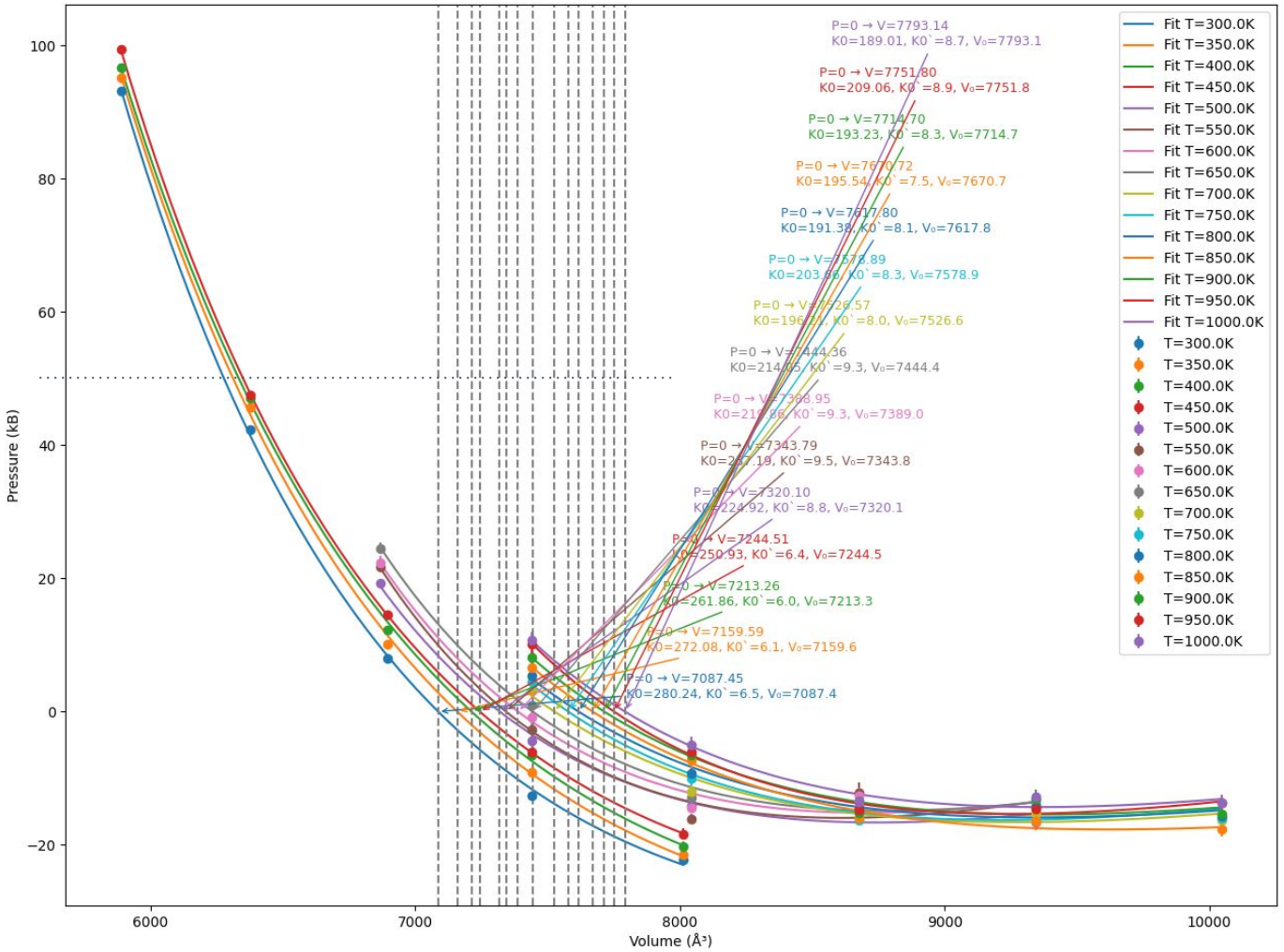
- Input: MD trajectory from VASP
- Process: Sampling > Structure Relaxation > Structural Analysis
- Output: Inherent Structures and CNA analysis (per atom / global)

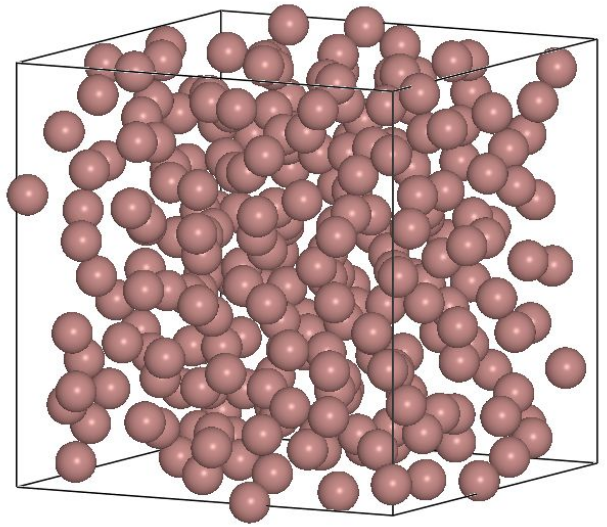
WF2 - Volume optimization

- Input: Collection of structures (like WF1 outputs) and temp. [list?]
- Process: Resize volume in range(70%,130%) > ~3ps NVTs > Thermodynamic
- Output: ExV and VxP plots, 1 bar predicted V



Workflows VASP - Module 2





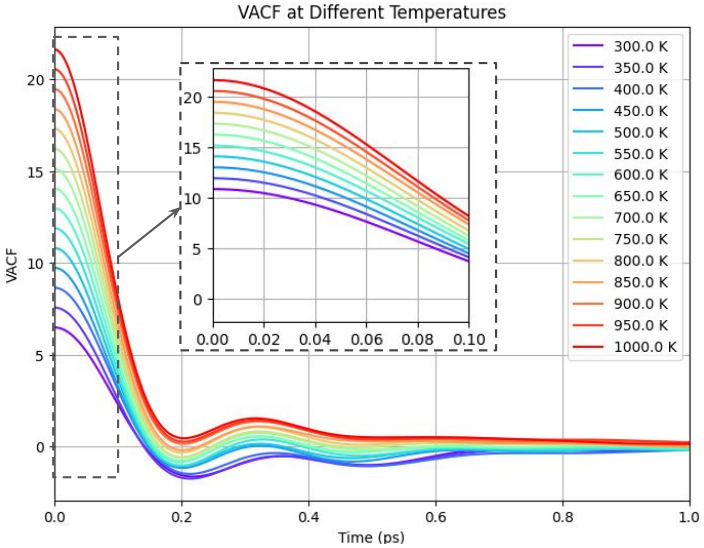
In_{250}

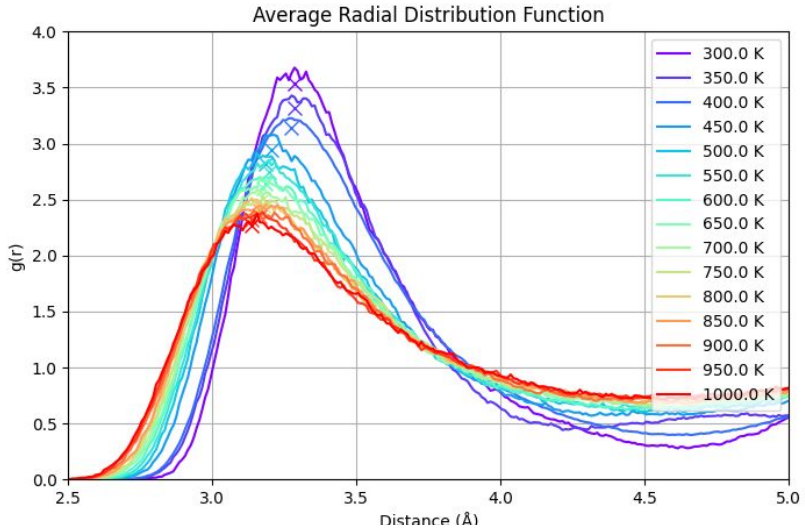
Each temperature studied:

5 *ab initio* MDs
64 cores/48h each

Volumes of equilibrium allow the study to be done at ideal thermodynamics conditions with standard NVT Molecular Dynamics!

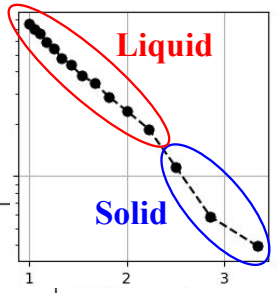
Workflows VASP - Module 2

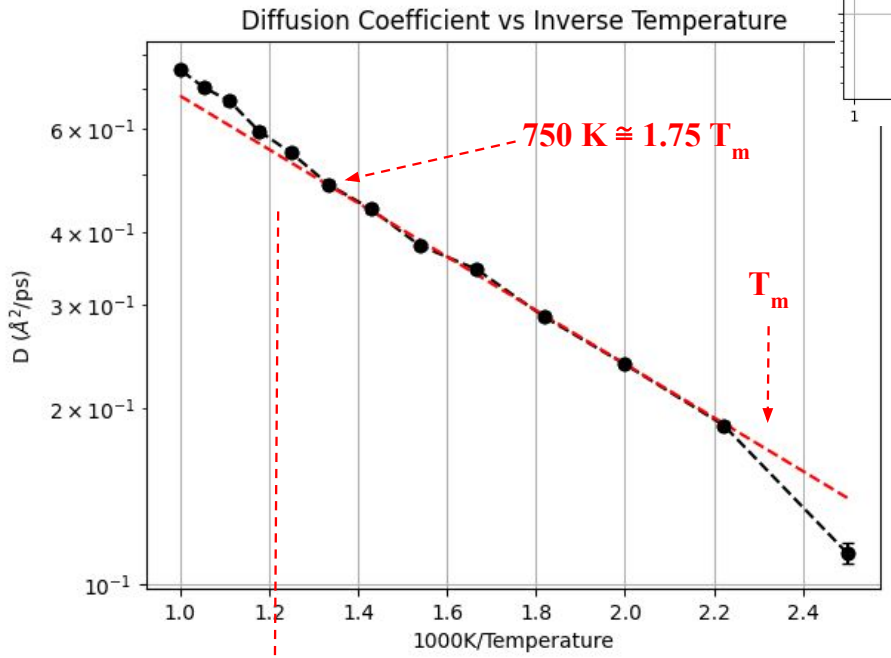




Indium melting point:

T_m=429.8 K





Arrhenius fit:
$D(x) = D_0 * \exp(s * x)$

D_0 = 1.949077 Å²/ps
s = -1.051576 (dimensionless per 'x' where x=1000/T)
=> $\ln(D) = \ln(D_0) - 1.051576 * (1000/T)$

Derived activation energy:

Ea = 0.090618 eV = 8.743 kJ/mol

Workflows VASP - Module 3

Workflows organization:

WF1 - MD structural analysis

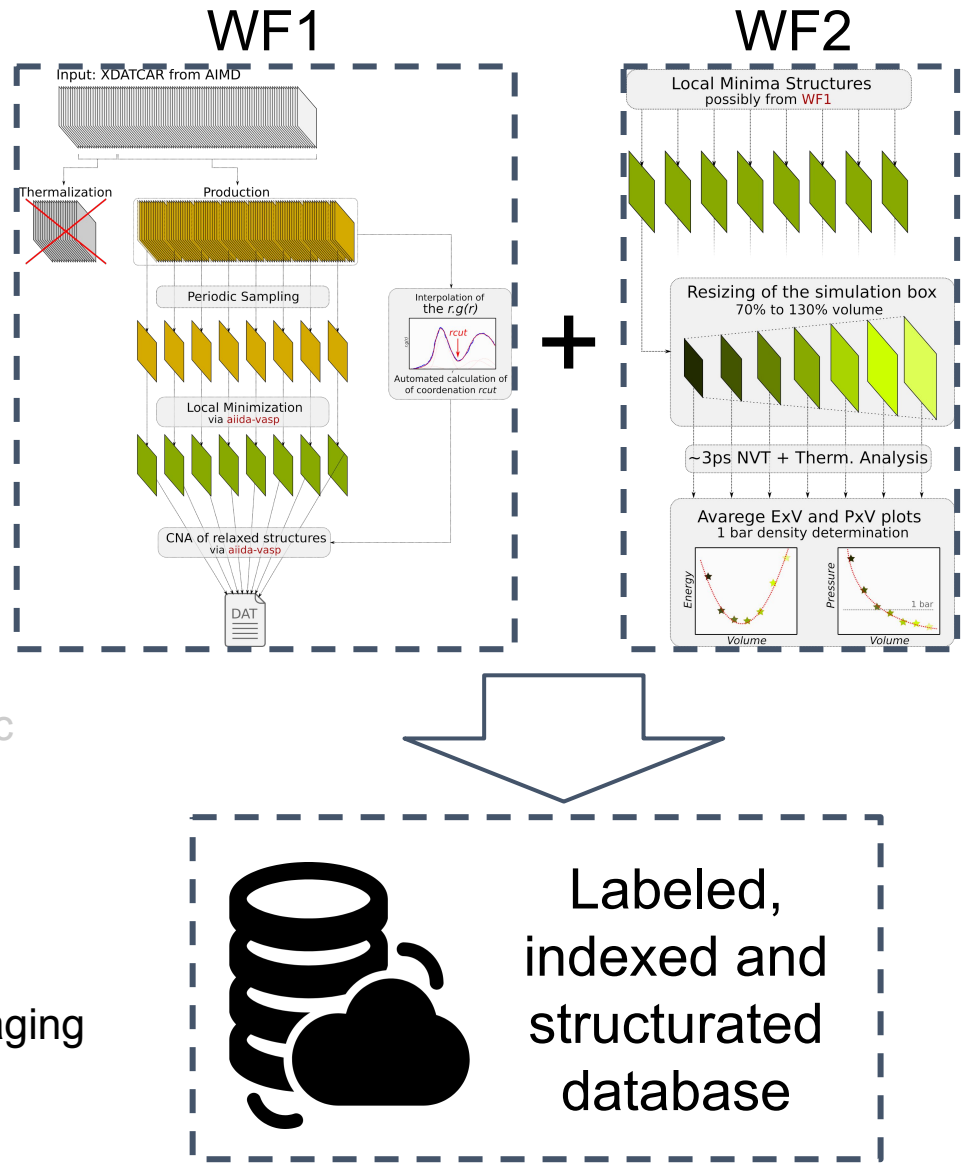
- Input: MD trajectory from VASP
- Process: Sampling > Structure Relaxation > Structural Analysis
- Output: Inherent Structures and CNA analysis (per atom / global)

WF2 - Volume optimization

- Input: Collection of structures (like WF1 outputs) and temp. [list?]
- Process: Resize volume in range(70%,130%) > ~3ps NVTs > Thermodynamic
- Output: ExV and VxP plots, 1 bar predicted V

WF3 - MLIP DataBase builder

- Input: Single structure, range of volumes and temperatures
- Process: Long NPT 1 bar > WF1 > WF2 > Single points / descriptors > Packaging
- Output: Database (structures, descriptors, energies, forces)

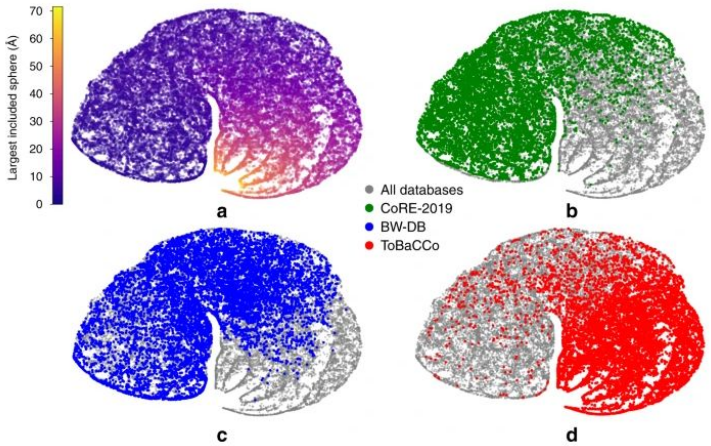


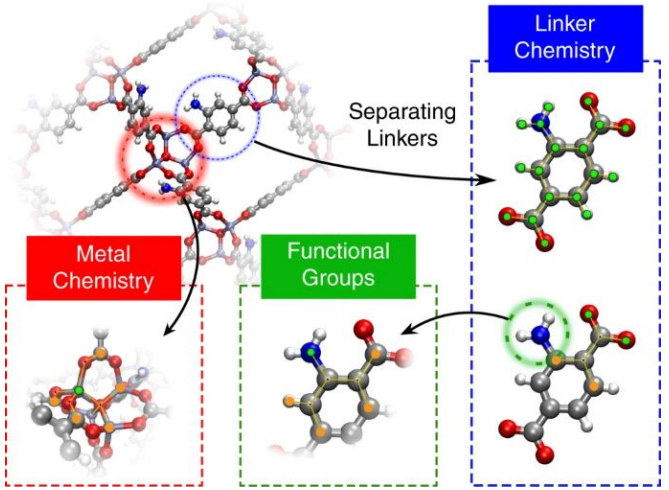
Use Case : MOF Learning

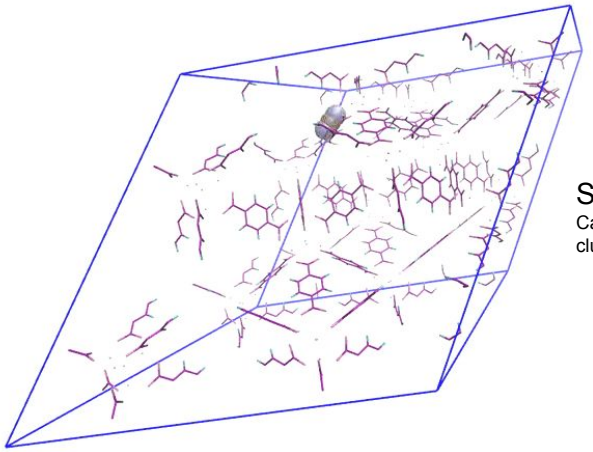
Workflow Python

- **Implementation and analysis:** A Hardiagon, FX Coudert
- **Technical support / containerization:** JPA de Mendonca, D Bissuel

- Compute **adsorption properties** (Monte-Carlo)
- Materials : Metal-organic Frameworks from **CoRE-MOF 2019 database**
- **single-CPU** calculations







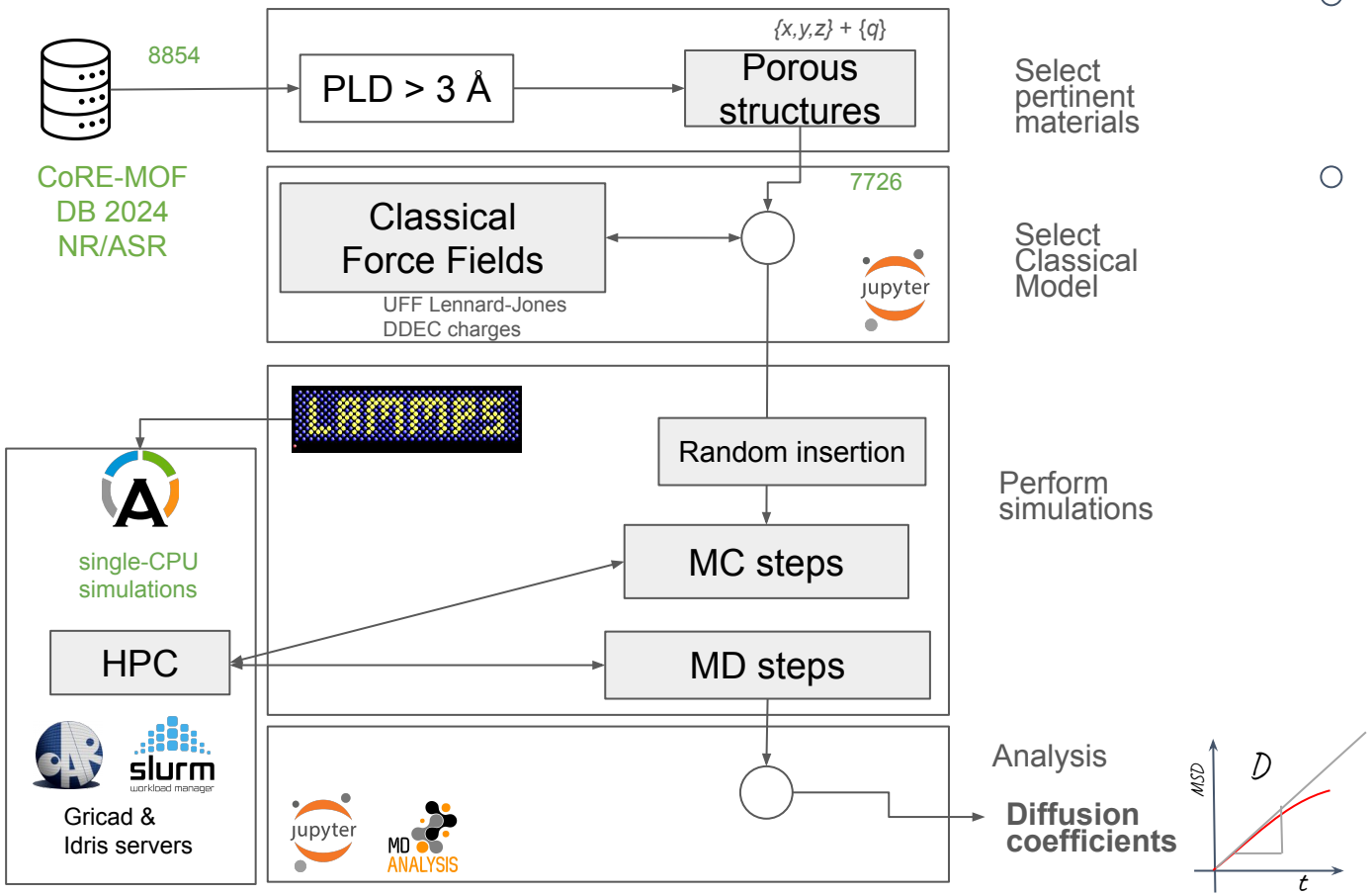
SBMOF-1
Ca-octahedra cluster+ SDB linker

Functionalities

- JSON input/outputs
- GUI for input/output analysis
- containers
- Benchmarks (Dahu, Gricad)

Use Case : MOF Learning

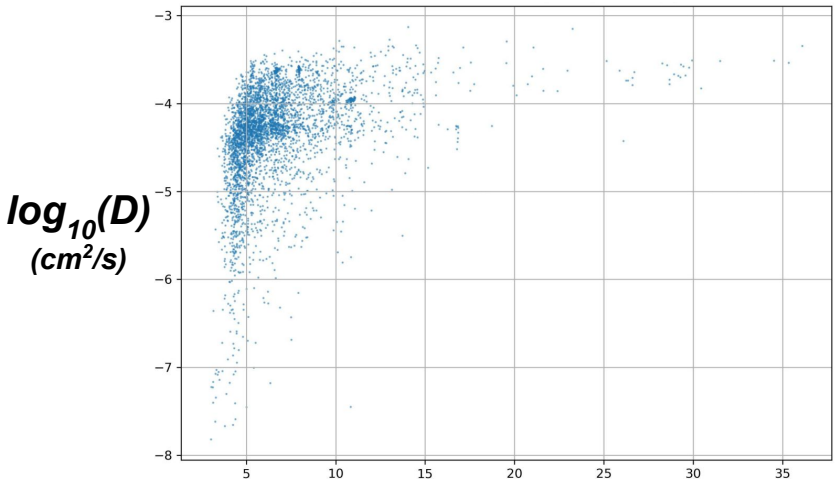
- **Workflow Diffusion**



- **Current State**
  - **Operational bash workflow**
    - use container with LAMMPS
    - screening with **Krypton** gas
  - **Aiida workflow tools**
    - simulations in 2 HPCs
    - store database
    - Template for Krypton
    - Template for CO_2

Intern student:
William Xu

2 weeks for 7726 1-cpu calculations (each ~48h)



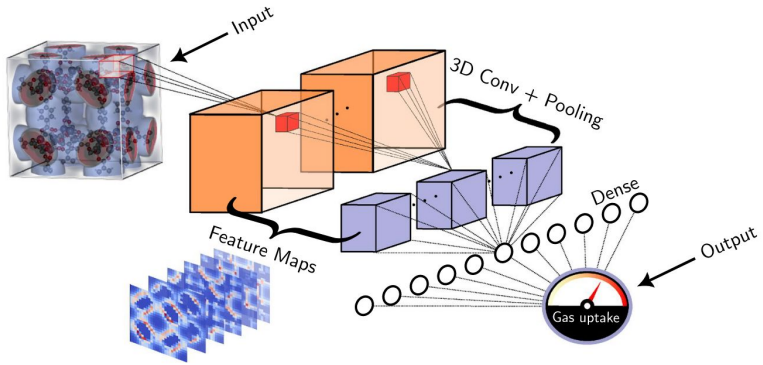
Use Case : MOF Learning

- **Deep Learning prediction**
  - **Small model :** Convolutional Neural Networks
  - **Fast predictions** 48h ⟶ 30 s
  - **Informative descriptor = energy grids (PES)**

$$\underbrace{x}_{\text{input}} \longrightarrow \underbrace{\phi(x;\theta)}_{\text{feature extraction}} \longrightarrow \underbrace{w^\top \phi(x;\theta) + w_0}_{\text{prediction}}$$

(fingerprint: $\phi(x;\theta)$; gas diffusion coefficient: $w^\top \phi(x;\theta) + w_0$)

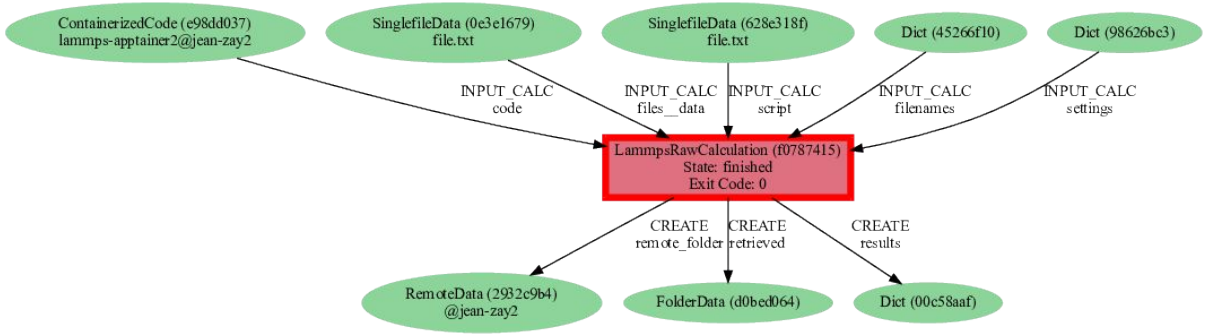
  - **Accuracy :** level of theory for PES evaluation



Sarikas, Scientific Reports, 2024

moxel

- **Perspectives**
  - **CNN training** on 3D maps
    - hyperparameters optimization
    - grid resolution
    - comparison with other methods (SOAP + RF)
  - **High-throughput** simulations with **CO_2** in MOFs
    - charges and electrostatics maps
    - Workflow manager : Aiida



Thanks!
Any questions?

https://diamond-diadem.github.io/en/