

FAIR research data with **NOMAD**

Operating NOMAD: A FAIR data service on the MPCDF

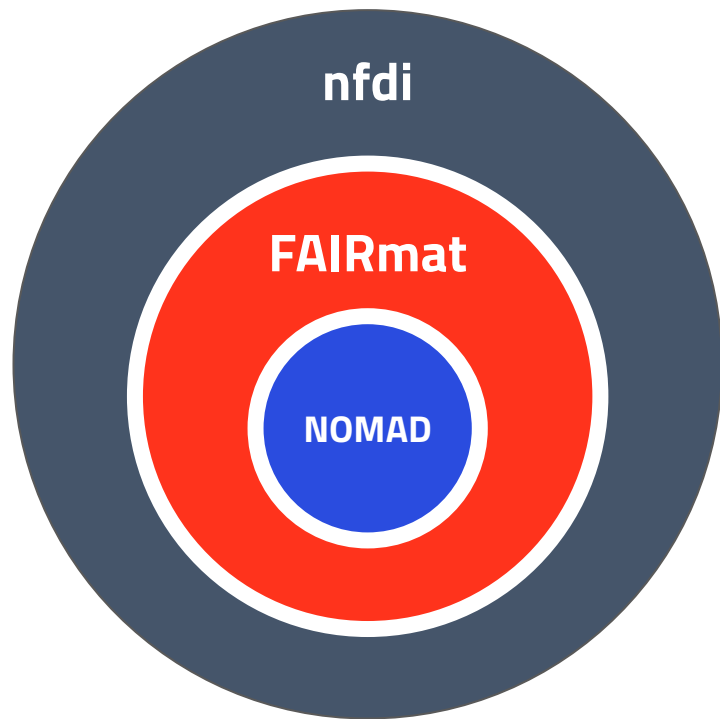
HPC Cloud with Kubernetes

Markus Scheidgen

Agenda

- ndfi, FAIRmat, and NOMAD
- Why kubernetes
- NOMAD at the HPC Cloud
- From commit to deployment
- Application monitoring

What are nfdi / FAIRmat / NOMAD

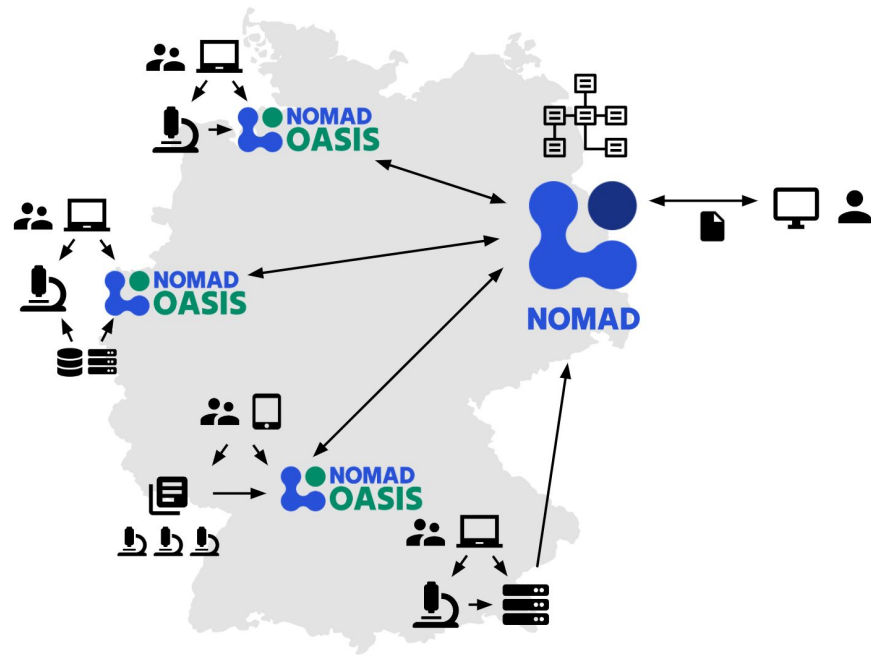
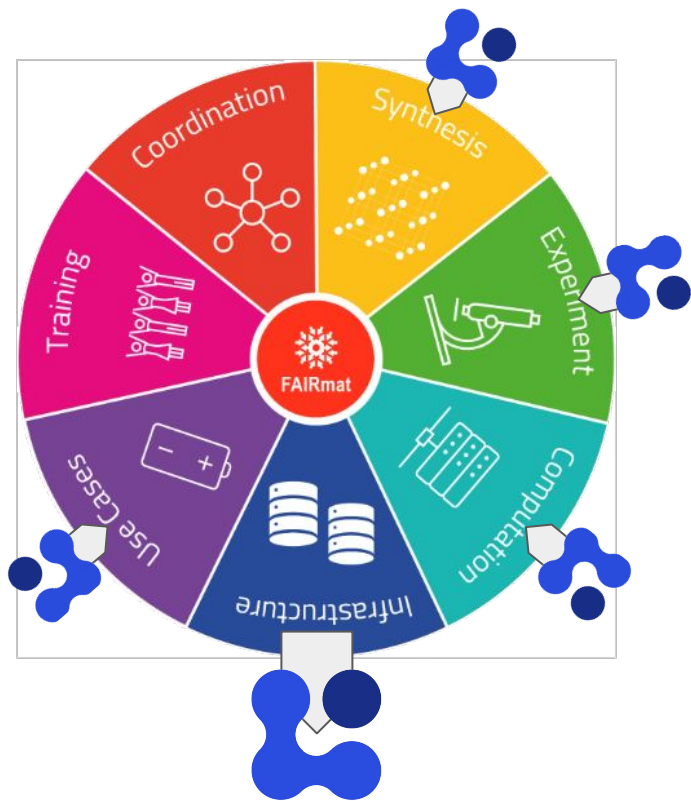


nfdi: Nationale Forschungsdaten Infrastructure, [link](#)
(national research data infrastructure)

FAIRmat: nfdi consortium for FAIR materials science data, [link](#)
(FAIR: findable, accessible, interoperable, re-usable)

NOMAD: a web-based service and software for managing FAIR materials science data, [link](#)
FAIRmat uses NOMAD to build a federated infrastructure of connected NOMAD installations

What are nfdi / FAIRmat / NOMAD



NOMAD

nomad-lab.eu/prod/v1/staging/gu/search/entries/entry/id/90f1Knlw01d2Et2tWncollm8b

PUBLISH EXPLORE ANALYZE ABOUT

Welcome Markus Scheidgen LOGOUT UNITS

Entries / Entry

OVERVIEW FILES DATA LOGS

Metadata

method name
DFT

program version
5.4.4.18Apr17-6-g9f103f2a35 (build 1...)

program name
VASP

basis set type
plane waves

core electron treatment
pseudopotential

Jacob's ladder
GGA

xc functional names
GGA_C_PBE, GGA_X_PBE

comment
no comment

references
matcloud.cnic.cn

authors
Zongguo Wang, Xushan Zhao et al

datasets
no datasets

mainfile
...17a4aed9067300267/vasprun.xml

entry id
90f1Knlw01d2Et2tWncollm8b

material id
sQqatLgYtaNbe4huwmDtJtXzGwV

upload id
t3xtCps5T9Wb7NEAV6ADGQ

upload create time
01/03/2018, 12:25:22

last processing time
08/01/2022, 06:08:52

processing version
1.0.0/foe3d31

Material

Original

Composition

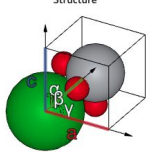
formula
BaO3Ti

dimensionality
bulk

elements
Ba, O, Ti

number of elements
3 (ternary)

Structure



Symmetry

crystal system
cubic

bravais lattice
cP

space group number
221

space group symbol
Pm-3m

point group
m-3m

structure name
unavailable

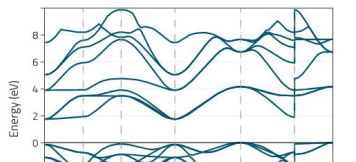
Lattice parameters

a	b	c
3.998 Å	3.998 Å	3.998 Å
α	β	γ
90 °	90 °	90 °
cell volume	63.926 Å ³	

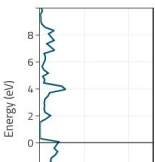
VIEW IN ENCYCLOPEDIA

Electronic properties

Band structure



Density of states



NOMAD

nomad-lab.eu/prod/v1/staging/gu/search/solarcells

PUBLISH EXPLORE ANALYZE ABOUT

Welcome Markus Scheidgen LOGOUT UNITS

Solar Cells

Filters

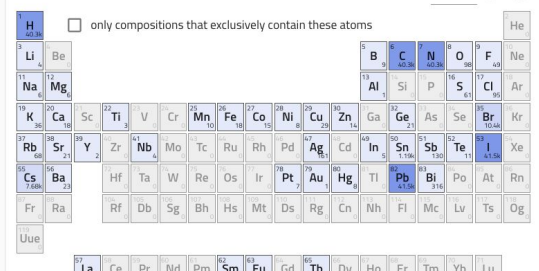
Type your query or keyword here

TERMS HISTOGRAM SCATTER PLOT PERIODIC TABLE

Elements

linear

only compositions that exclusively contain these atoms



Solar Cell Absorber Fabricati...

linear

Type here

- Spin-coating 29.4k
- Spin-coating >> Spin-coating 5.49k

SHOWING TOP 2 ITEMS

Solar Cell Device Stack

linear

Type here

- Perovskite 42.6k
- SLG 4.16k

SHOWING TOP 2 ITEMS

Scatter plot

Efficiency

Open Circuit Voltage (V)

Scatter plot

Efficiency

Open Circuit Voltage (V)

Solar Cell Electron... linear

Solar Cell Hole Tra... linear

Band Gap Value (eV)

25k

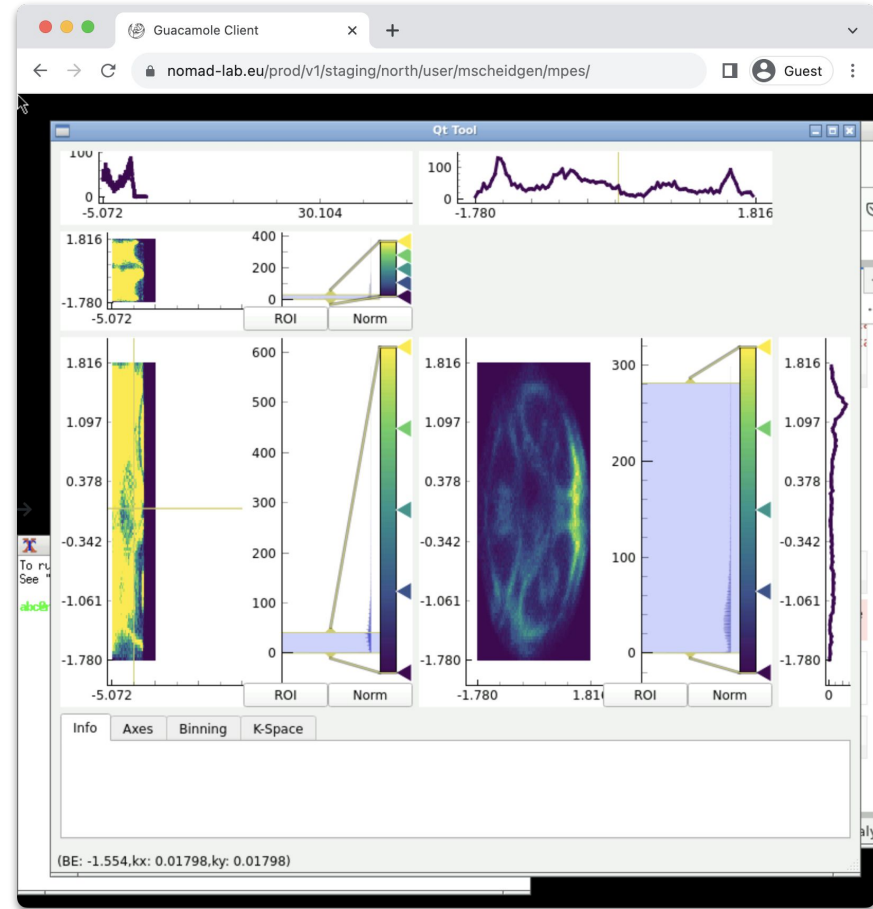
The screenshot shows a JupyterLab environment with a file browser on the left and a code editor in the center. The code defines an ArchiveQuery and a workflow for calculating energy values.

```

from nomad.client import ArchiveQuery
from nomad.metainfo import units

query = ArchiveQuery(
    query={
        'results.method.simulation.program_name': 'VASP',
        'results.material.elements': ['Ti', 'O'],
        'results.properties.geometry_optimization': {
            'final_energy_difference': {'lt': 1e-22,
        }
    },
    required={
        'workflow': {
            'calculation_result_ref': {
                'energy': '*',
                'system_ref': {
                    'chemical_composition_reduced': '*'
                }
            }
        }
    }
})

for result in query.download(100):
    calc = result.workflow[0].calculation_result_ref
    formula = calc.system_ref.chemical_composition_reduced
    if calc.energy.total:
        total_energy = calc.energy.total.value.to(units.eV)
    else:
        total_energy = 'N/A'
    print(f'{formula}: {total_energy}')
  
```



NOMAD

nomad-lab.eu/prod/v1/gui/search/entries/entry/id/sCItplLugL04XS2EVLsv6Us6Eznlj/data/run/system/atoms/positions

Welcome Markus Scheidgen LOGOUT UNITS

PUBLISH EXPLORE ANALYZE ABOUT

Entries / Entry / Data

OVERVIEW FILES DATA LOGS

search

code specific all defined definitions

Entry

section

SUB SECTIONS

- results
- metadata
- workflow
- run**

REFERENCED BY [closed](#)

Run

section

SUB SECTIONS

- program
- method
- system
- calculation**

REFERENCED BY [closed](#)

System

section

QUANTITIES

- type = bulk
- configuration_raw_gid = gSWzj-w8k853xs3i_x34kF6VPcJU
- is_representative = true
- chemical_composition = BaO00Ti
- chemical_composition_hill = BaO3Ti
- chemical_composition_reduced = BaO3Ti

SUB SECTIONS

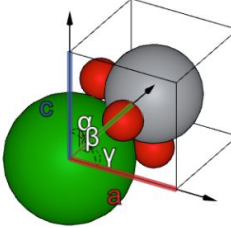
- atoms**
- springer_material
- symmetry

REFERENCED BY [closed](#)

Atoms

section

- O
- Ti
- Ba



QUANTITIES

- species = 5 vector
- labels = 5 list
- positions = 5 x 3 matrix**
- lattice_vectors = 3 x 3 matrix
- periodic = 3 list

REFERENCED BY [closed](#)

positions

quantity

VALUE

0	0	2
0	2	0
0	2	2

(n_atoms x 3)
Å

REFERENCED BY [closed](#)

NOMAD pro

keycloak

homepage

(raw) files

vasp.run.xml

INCAR

OUTCAR

experimenter

snapshot.png

sample.json

processed data

task queue

- system
- calculation (DOS)
- more (meta)data

worker

- instrument
- sample
- data
- more (meta)data

aggregations

jupyterhub

- material
- system
- properties

notebook

- thermal
- more (meta)data

app

portal

exploration
domain specific

load balancer

tools
pre installed tools
and programming
environments

GUI

API

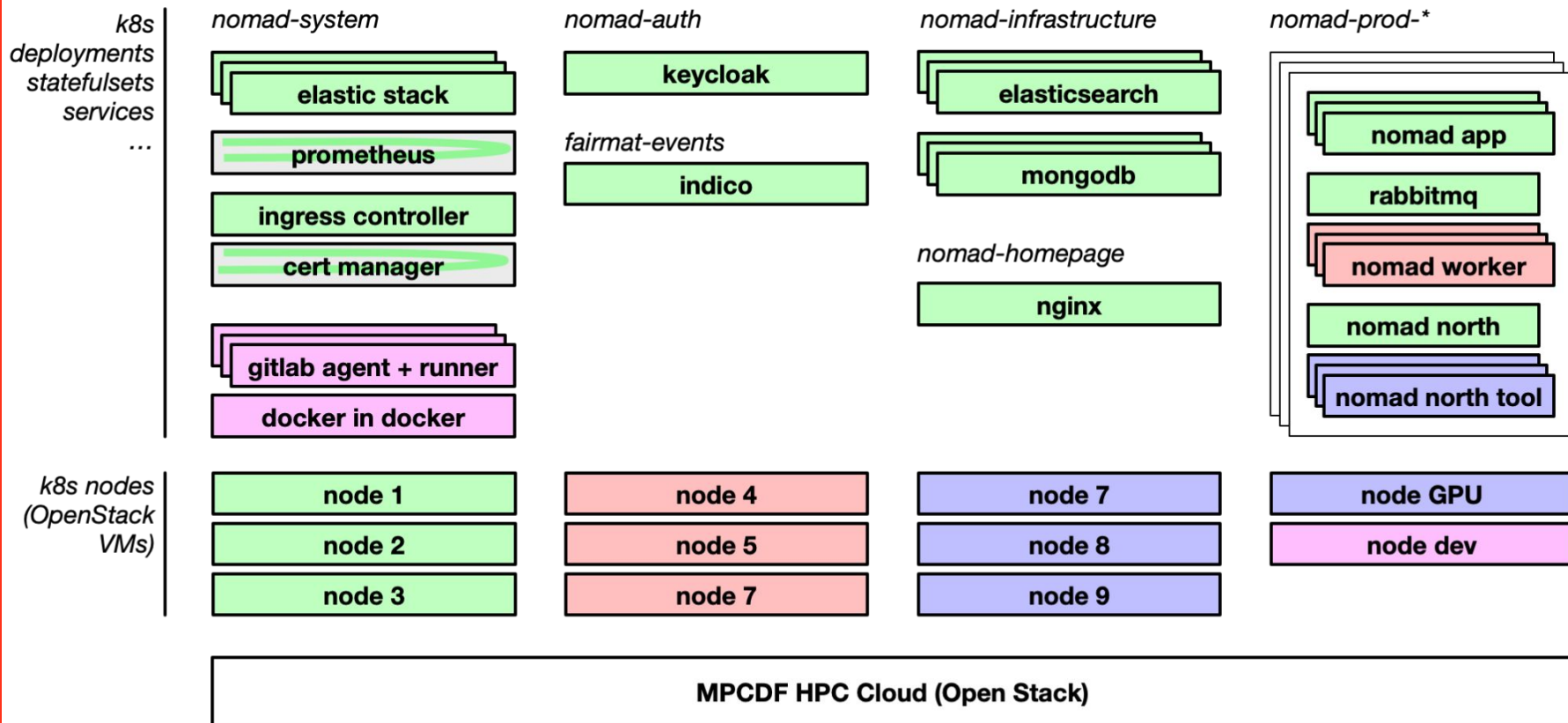
mongodb

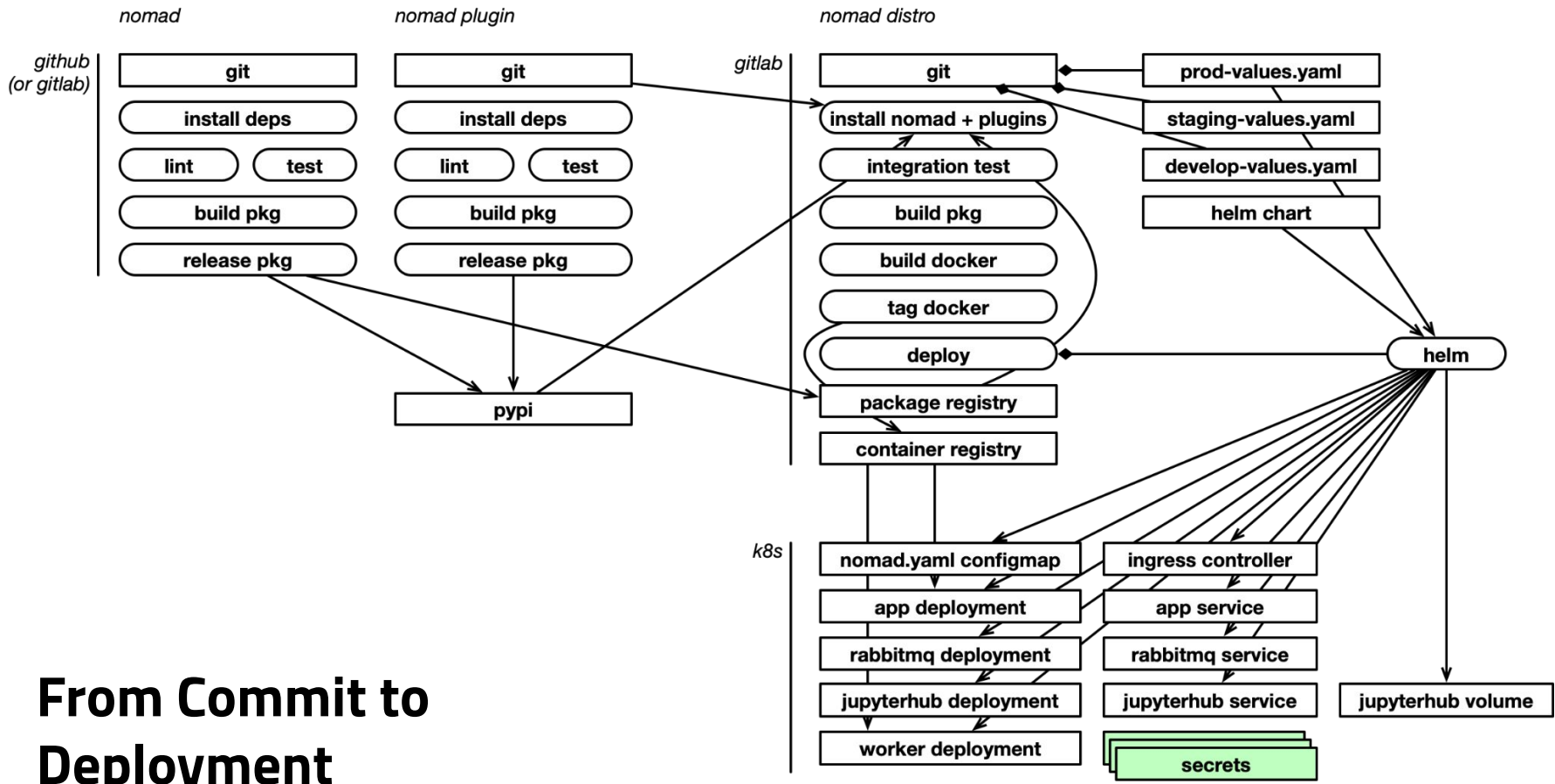
formal schema
Metainfo

elasticsearch



NOMAD on the HPC Cloud

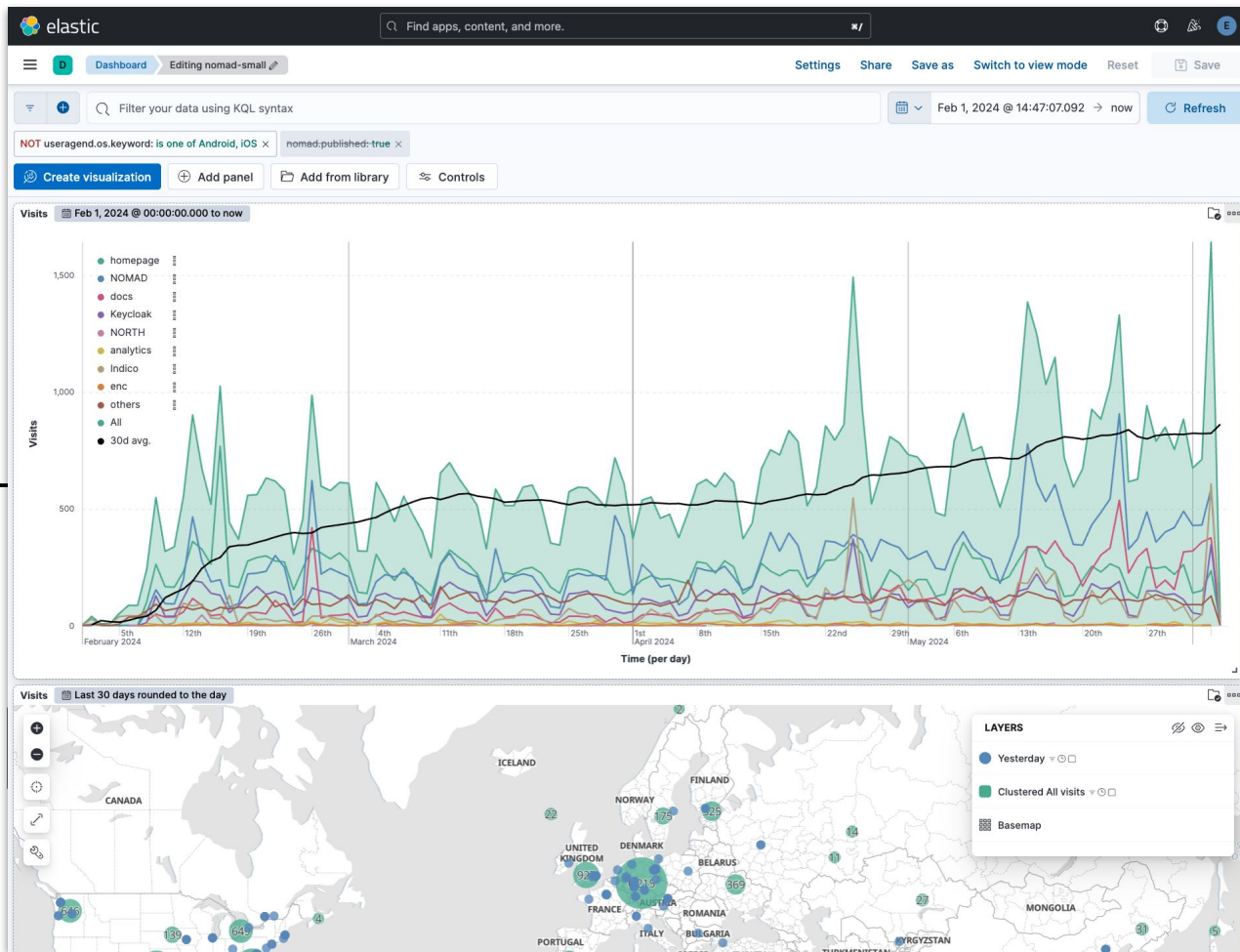




From Commit to Deployment

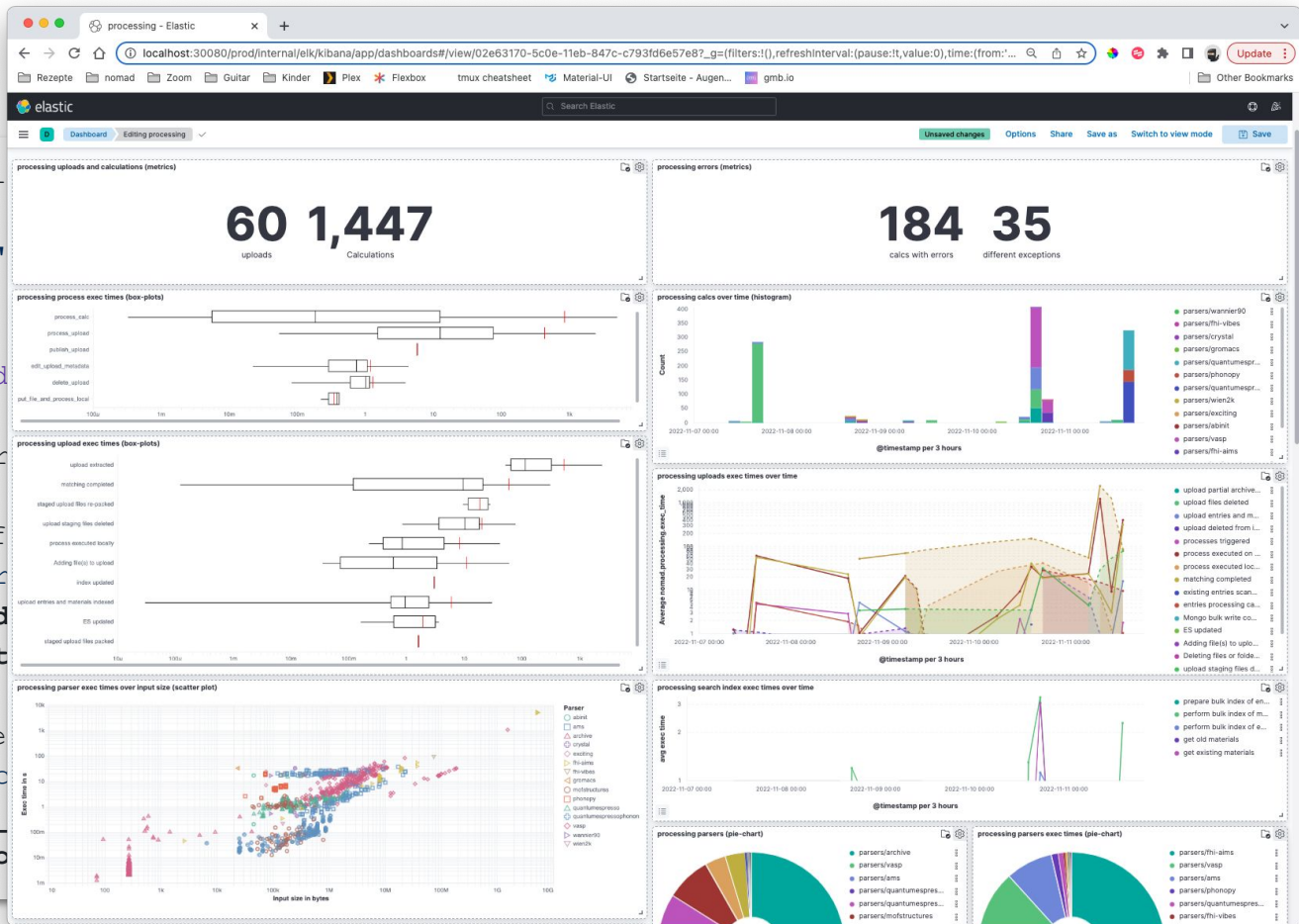
Elasticstack

keycloak



Structlog

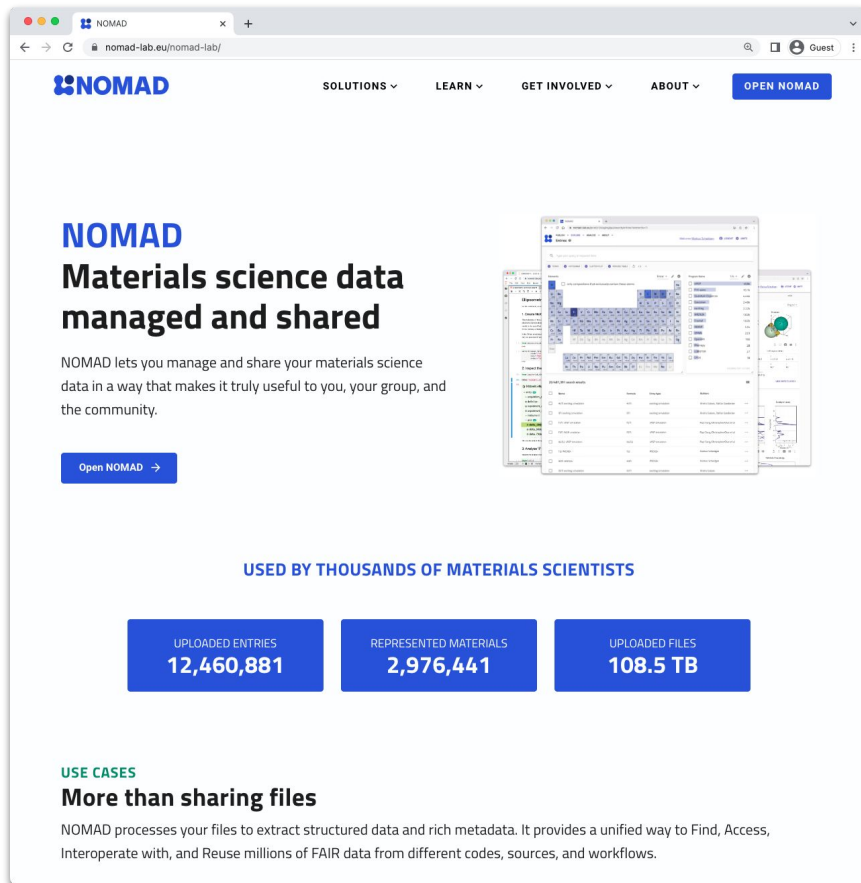
```
1. logger = get_logger(__name__,  
2.     __name__,  
3.     api="upload")  
4.  
5. @app.get  
6. def get_upload_data():  
7.     try:  
8.         with timer:  
9.             ...  
10.            logger.info("uploading")  
11.            return upload_data()  
12.            upload_data() exec_time()  
13.            exec_time()  
14.        except e:  
15.            logger.error("uploading failed")  
16.            exc_info = sys.exc_info()  
17.            exc_info = exc_info[0]  
18.            upload_data() exec_time()
```



Summary

- “easy” orchestration and scaling to create a complex service
- automated pipelines from commit to deployment
- monitoring an application on millions of logs

Questions?



NOMAD
Materials science data managed and shared

NOMAD lets you manage and share your materials science data in a way that makes it truly useful to you, your group, and the community.

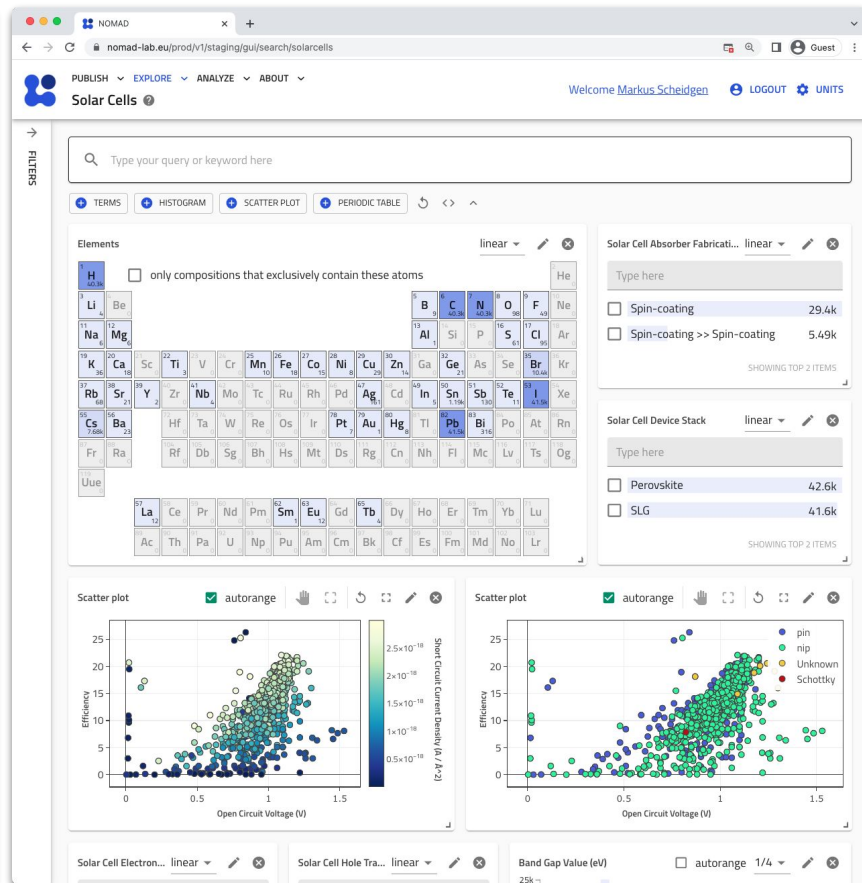
[Open NOMAD](#)

USED BY THOUSANDS OF MATERIALS SCIENTISTS

UPLOADED ENTRIES 12,460,881	REPRESENTED MATERIALS 2,976,441	UPLOADED FILES 108.5 TB
---------------------------------------	-------------------------------------------	-----------------------------------

USE CASES
More than sharing files

NOMAD processes your files to extract structured data and rich metadata. It provides a unified way to Find, Access, Interoperate with, and Reuse millions of FAIR data from different codes, sources, and workflows.



Solar Cells

SEARCH: Type your query or keyword here

Filters: TERMS, HISTOGRAM, SCATTER PLOT, PERIODIC TABLE

Elements: only compositions that exclusively contain these atoms

<input type="checkbox"/> Spin-coating	29.4k
<input type="checkbox"/> Spin-coating >> Spin-coating	5.49k

SHOWING TOP 2 ITEMS

<input type="checkbox"/> Perovskite	42.6k
<input type="checkbox"/> SLG	41.6k

SHOWING TOP 2 ITEMS

Scatter plot 1: Efficiency vs. Open Circuit Voltage (V). Legend: autorange. Y-axis: 0 to 25 . X-axis: 0 to 1.5 . Color scale: Short Circuit Current Density (A/A^2) from 0.5×10^{-18} to 2.5×10^{-18} .

Scatter plot 2: Efficiency vs. Open Circuit Voltage (V). Legend: autorange. Y-axis: 0 to 25 . X-axis: 0 to 1.5 . Legend items: pin (blue), nip (green), Unknown (yellow), Schottky (red).

Band Gap Value (eV): autorange 1/4

NOMAD Links

- homepage – nomad-lab.eu
- central NOMAD service – <https://nomad-lab.eu/prod/v1>
- [NOMAD Dokumentation](#)
 - [installing Oasis](#)
 - [schemas and ELNs](#)
- Tutorials and videos
 - [NOMAD Oasis tutorial](#) video playlist
 - [Publishing, Exploring, API](#) video playlist
 - All FAIRmat [tutorials](#)
- [NOMAD forum](#)
- NOMAD's [main gitlab](#) (and [github](#))