NOMAD and NOMAD Oasis

A FAIR Data-Management Platform for Materials Science

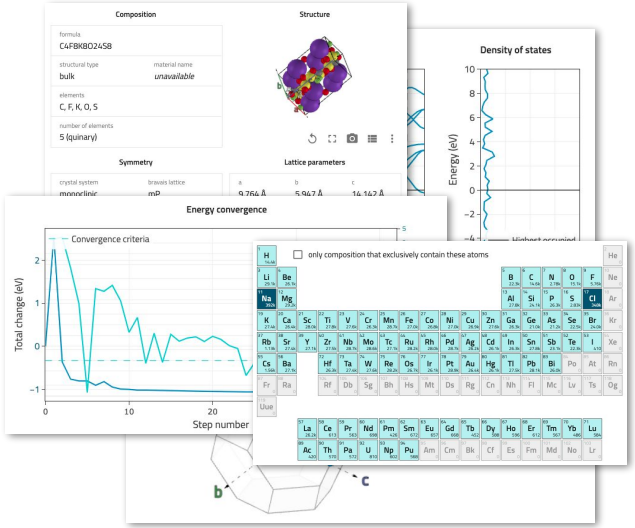
Markus Scheidgen, Sandor Brockhauser

Meet the team

NOMAD: Publishing research data

More than 12 million of simulations (22 billion quantities) from over 500 authors world-wide

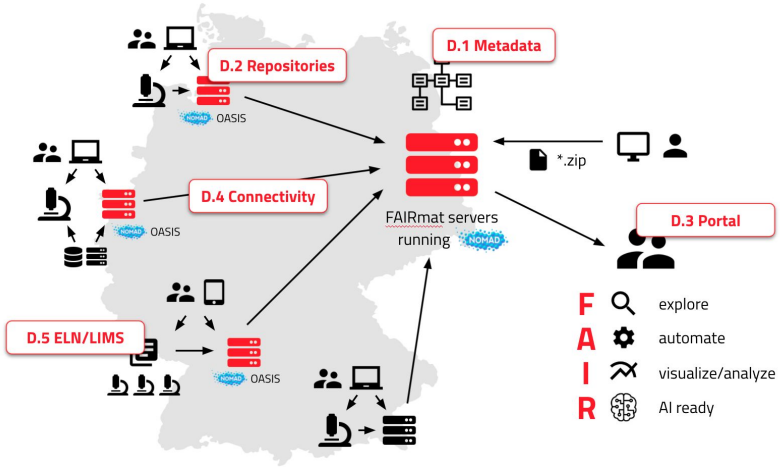
- ➔ Free publication and sharing data of data
- ➔ Extracts **rich metadata** for more than **50 codes**
- ➔ All data in a **raw** and a common **machine readable** from
- ➔ Use integrated tools to **explore**, **visualize**, and **analyze**



Oasis: Local data management

FAIRmat develops and uses NOMAD software to build a federated infrastructure of local repositories

- ➔ Organise research data though its whole life-cycle
- ➔ Adaptable to your workflows and data-types
- ➔ A first step to connect with in the FAIRmat network
- ➔ Oasis is being developed and you can shape its future



PUBLISH EXPLORE ANALYZE ABOUT

Search › Entry

LOGIN / REGISTER

jupyterhub query_nomad_archive (unsaved changes)

Logout Control Panel

File Edit View Insert Cell Kernel Widgets Help

Trusted Python 3

Run Markdown

Querying the Archive and performing Artificial Intelligence modeling

created by: Luigi Sbailò,[1] Matthias Scheffler,[1] and Luca Ghiringhelli[1]

[1] Fritz Haber Institute of the Max Planck Society, Faradayweg 4-6, D-14195 Berlin, Germany

ghiringhelli@fhi-berlin.mpg.de, sbailo@fhi-berlin.mpg.de

[Last updated: Jan 5, 2021]

MAX-PLANCK-GESELLSCHAFT

NOMAD NOVEL MATERIALS DISCOVERY

In this tutorial, we show how to query the NOMAD Archive (https://www.nomad-coe.eu/index.php?page=nomad-repository) and perform data analysis on the retrieved data.

Preliminary operations

We load the following packages that are all available in the virtual environment containing the Jupyter notebooks in the NOMAD AI toolkit. Among the loaded packages, we highlight `sklearn`, i.e., scikit-learn, one of the most popular machine-learning packages, and `pandas`, a popular tool for data handling and analysis.

In [ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import os
import math
import plotly.express as px
from tqdm import tqdm
from sklearn import preprocessing, tree
from sklearn.manifold import TSNE
from sklearn.model_selection import train_test_split
from query_nomad_archive.visualizer import Visualizer
from sklearn import decomposition
from IPython.display import display, Markdown
import re
```
executed in 1.53s, finished 17:29:39 2021-04-14

In [ ]:
```
from nomad import client, config
```
executed in 977ms, finished 17:29:41 2021-04-14

In [ ]:
```
from nomad.client import ArchiveQuery
from nomad.metainfo import units
```
executed in 2ms, finished 17:29:41 2021-04-14

We maintain a `nomad` package that can be imported in all notebooks of the AI Toolkit.

The `nomad` package allows to retrieve data from the NOMAD Archive by means of a script, as shown below. In this script, we insert metadata characterizing the materials that we aim to retrieve. In this case, we select ternary materials containing oxygen. We also request that simulations were carried out using the VASP code using GGA exchange-correlation (xc) functionals. Values are retrieved from the simulation run that found geometrically convergence within a desird threshold value of.

Guacamole Client

localhost:9000/fairdi/nomad/latest/north/user/mscheidg/pyarpes/

Update

pyARPES_down... - JupyterLab — Mozilla Firefox

pyARPES_down... - Jupyt

Qt Tool

File Edit

Launcher

Normalization

Operation: Divide Subtract Off

Mean: ROI Frame Time range

Blur: X 0.00 Y 0.00 T 0.00

Info Axes Binning K-Space

(BE: -1.554,kx: 0.01798,ky: 0.01798)

Simple 1 1 Python 3 (ipykernel) | Busy Mode: Command Ln 1, Col 1 pyARPES_download_and_visualize.ipynb

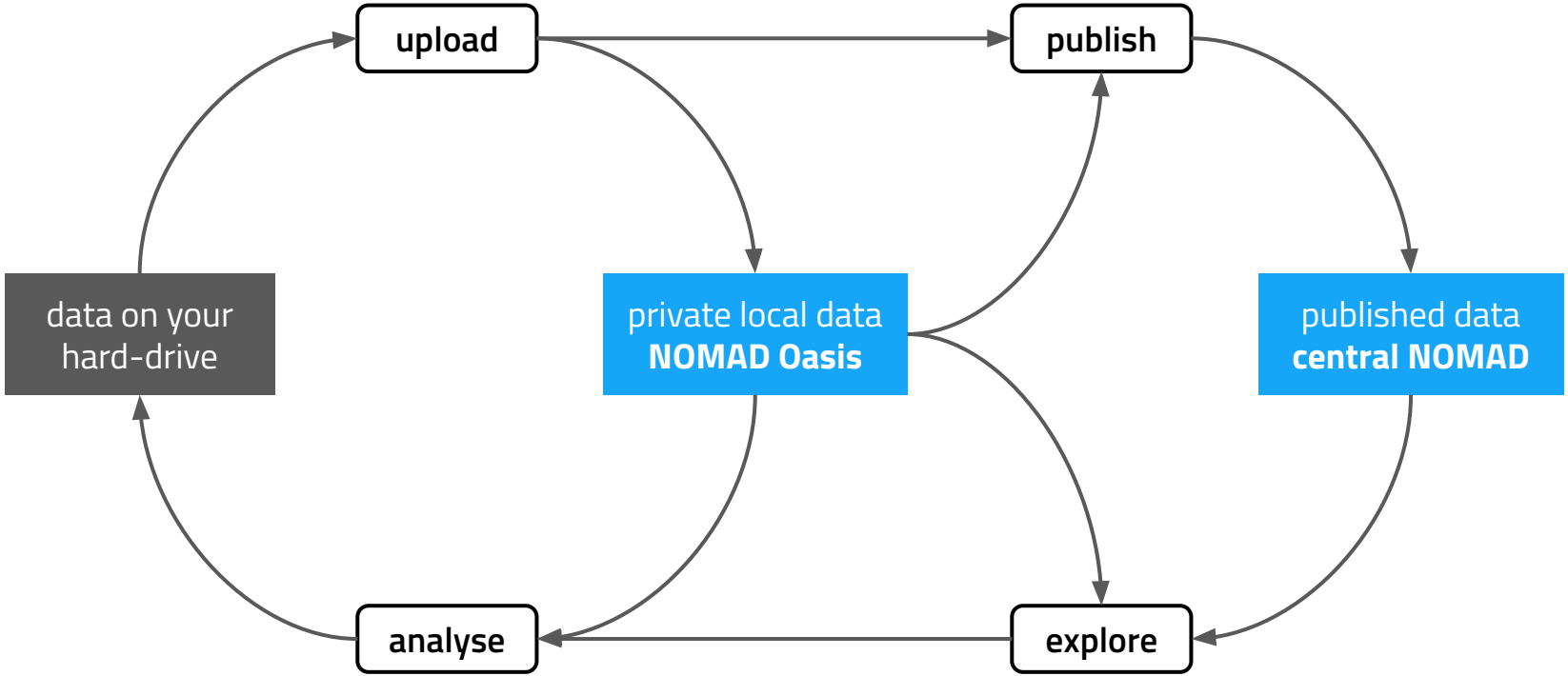
ROI Norm

UNITS

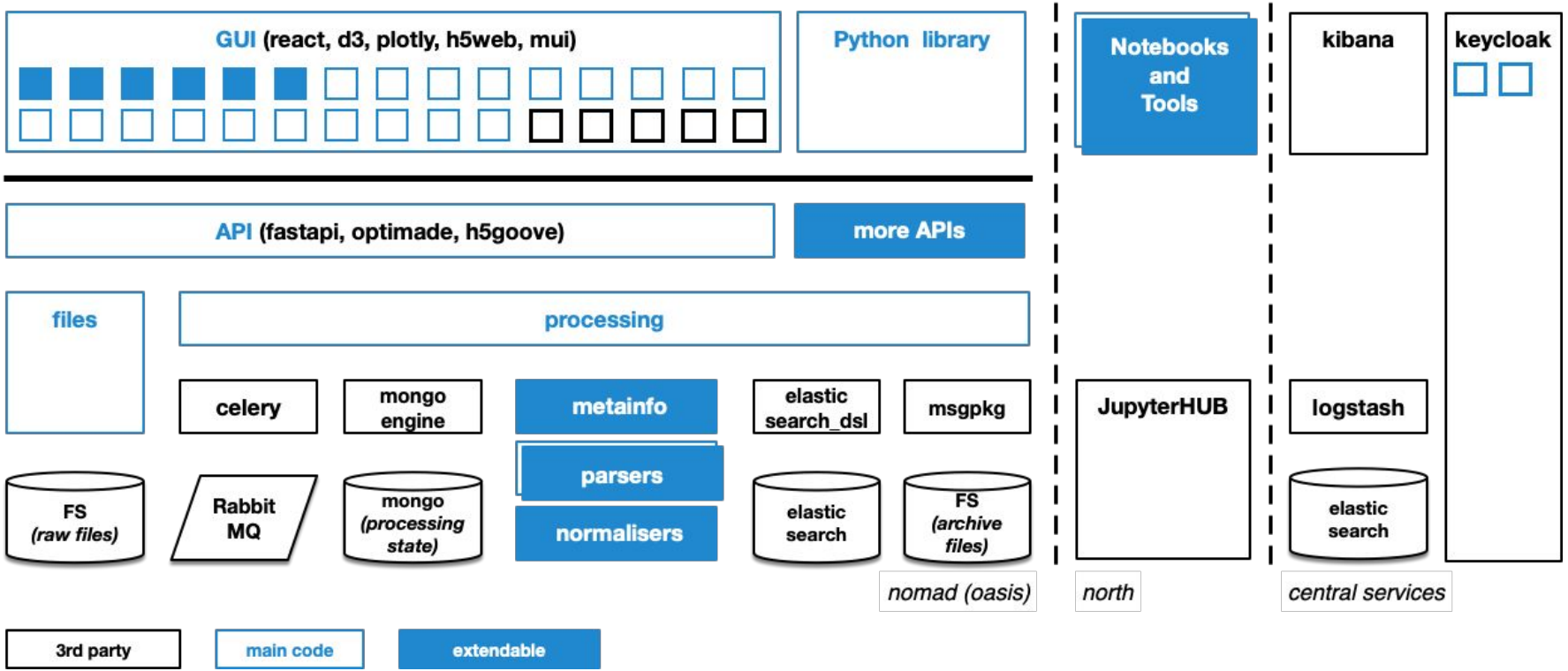
6000

305.57120

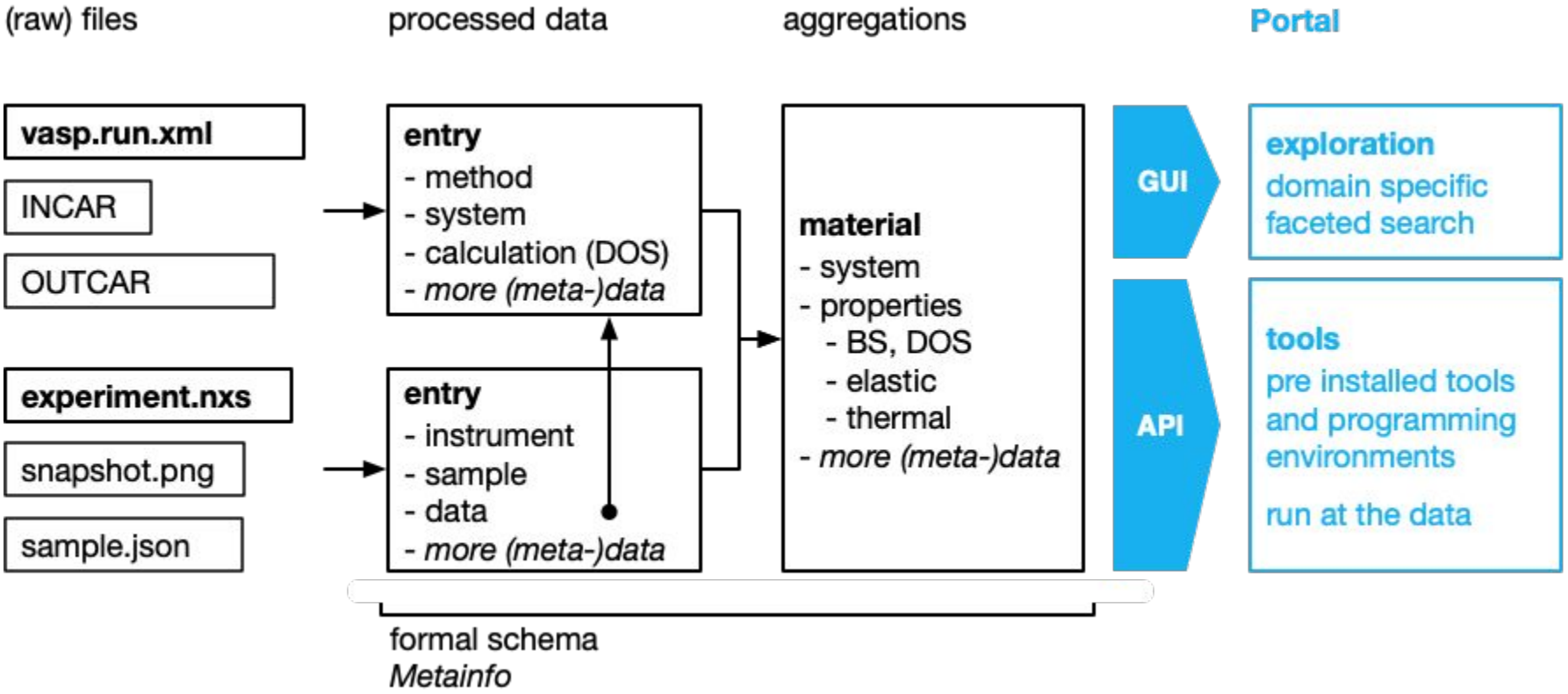
NOMAD to assist research processes



Architecture



Core functionality: Processing data files to extract (meta-)data

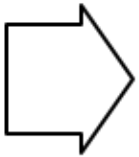


Extended functionality

ELN | NORTH | neXus | parsers | portal

api
uploading, search, download, partial-access

raw files
uploaded files "as they are uploaded"



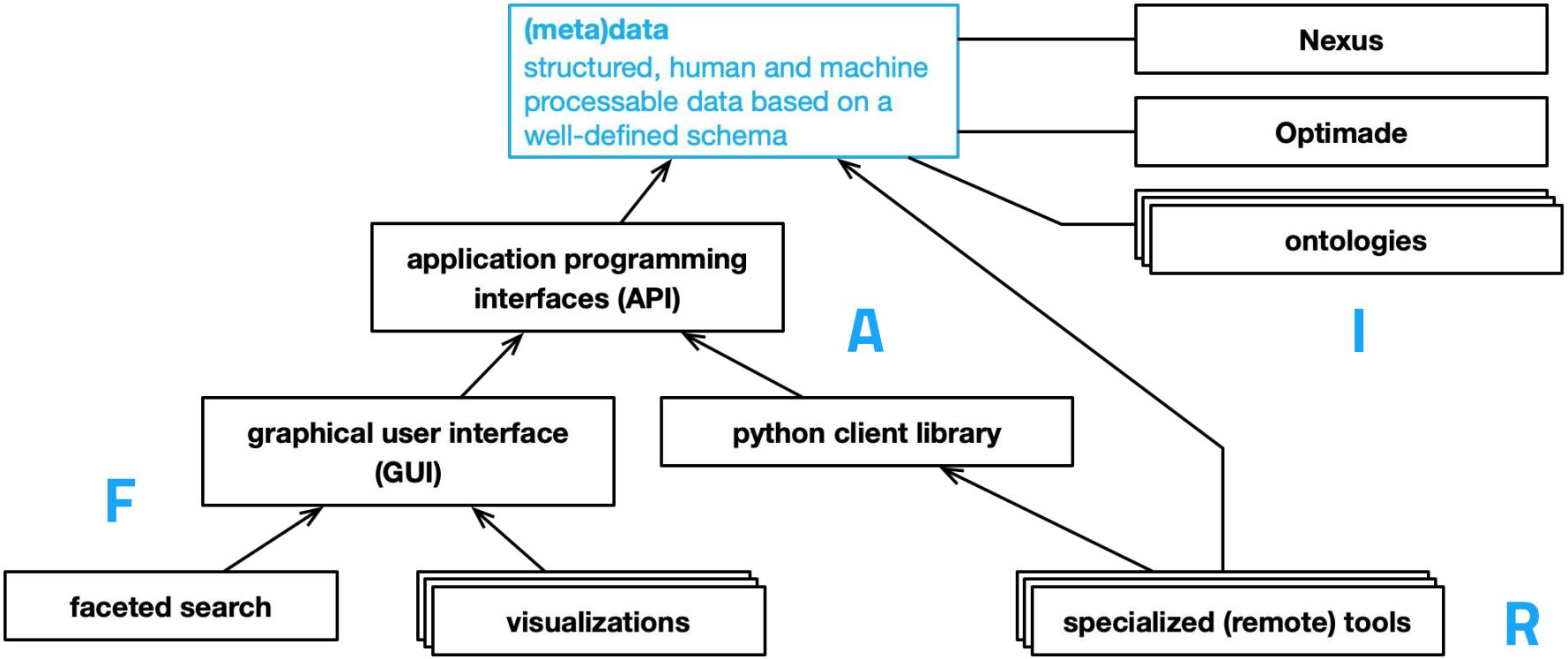
processed (meta-)data
structured schema-based (*metainfo*) data documents

uploaded files

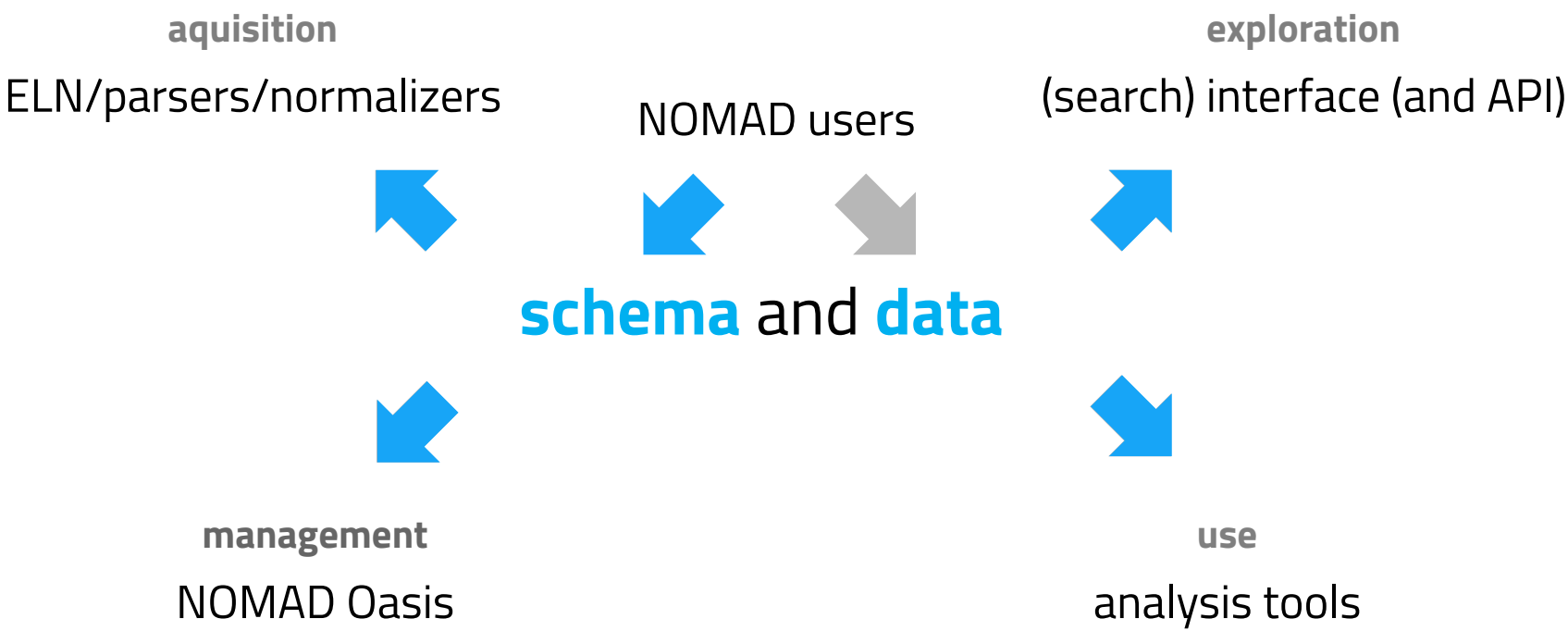
binary files | **mongo db** | **elastic search**

central operation | oasis installation | oasis network | data transfer

Well-defined schemas make data FAIR

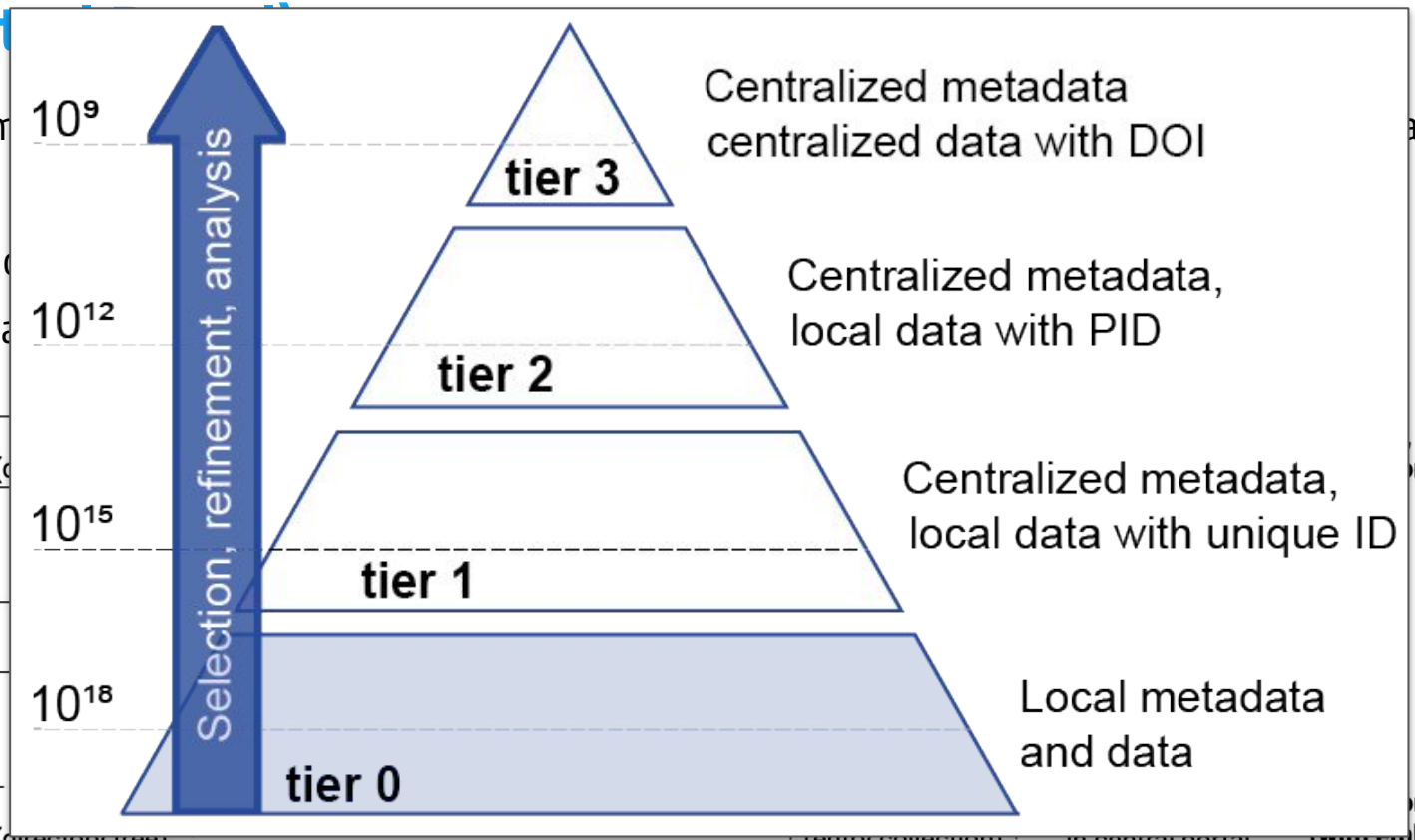


Custom schemas



Oasis data transfer: From Installation (e.g. Oasis) to Installation (e.g. Cent

- NOMAD m
  collection
- These upl
- *Push* uploa



users
or software

(directory tree)

(entry collection)

in central portal

and a

r "public"

r published
bargo)

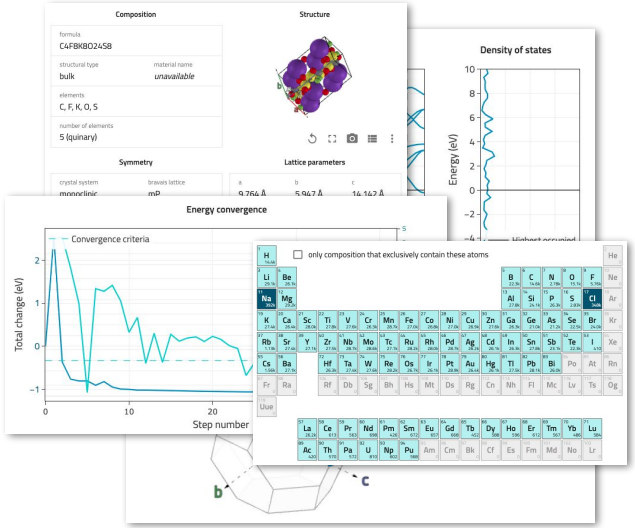
Conclusions

- bottom up towards FAIR data
- a flexible, extendable, schema-focused framework for research data management in materials science
- data sharing and exploration in a federated network of data repositories

NOMAD: Publishing research data

More than 12 million of simulations (22 billion quantities) from over 500 authors world-wide

- ➔ Free publication and sharing data of data
- ➔ Extracts **rich metadata** for more than **50 codes**
- ➔ All data in a **raw** and a common **machine readable** from
- ➔ Use integrated tools to **explore**, **visualize**, and **analyze**



Oasis: Local data management

FAIRmat develops and uses NOMAD software to build a federated infrastructure of local repositories

- ➔ Organise research data though its whole life-cycle
- ➔ Adaptable to your workflows and data-types
- ➔ A first step to connect with in the FAIRmat network
- ➔ Oasis is being developed and you can shape its future