The NOMAD Encyclopedia

Georg Huhs

Getting access to millions of calculated materials properties

This project has received funding from the European Union's Horizon 2020 research and innovation programme, grant agreement No 676580.

Context

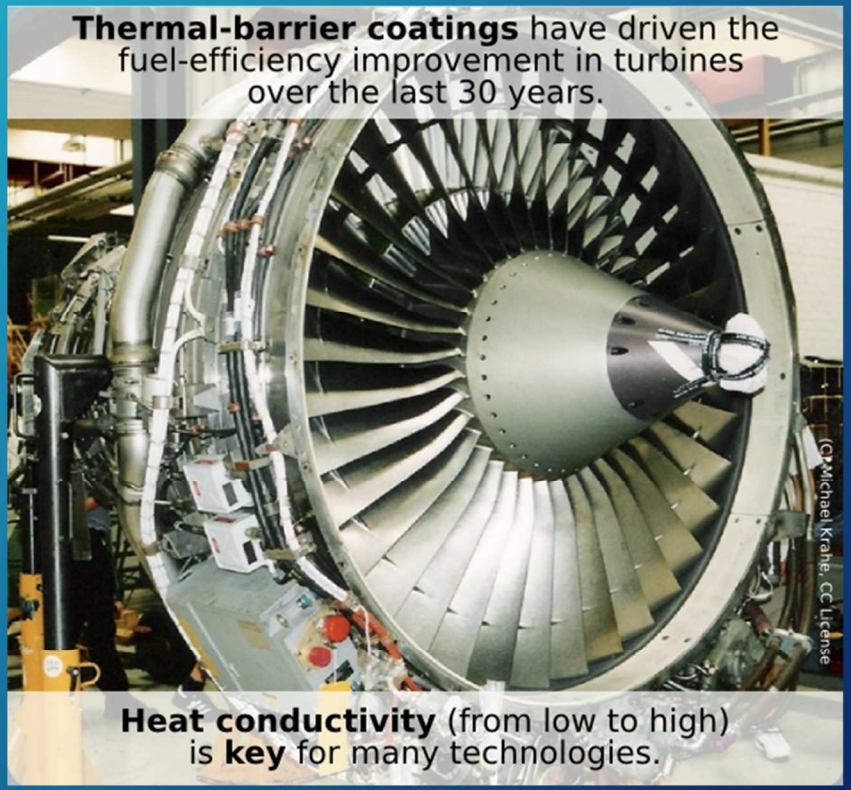
European Centre of Excellence (H2020)

- Provide services to industry
- User-driven
- Based on high performance computing

Field: material science **DATA**

Currently: Computational (experimental data is long term objective)

Ab-Initio (classical MD in progress, more to come)



Materials data and their structure

NOMAD

Level	Properties	Methods	Size
I	Atomic positions and nuclear charges, properties of free atoms, symmetry, temperature, pressure	Input: definition of material *gene*	10 kB - 10 MB
II	Tota wav geo		10 MB - 10 TB
III	Exc mat opti pho		1 GB - 1 TB
IV	Efficiency of solar cell, thermoelectric figure of merit, turn-over frequency of catalyst, etc. as a function of temperature and pressure	Modeling, output derived from levels I-III *phenotype*	10 kB - 1 MB

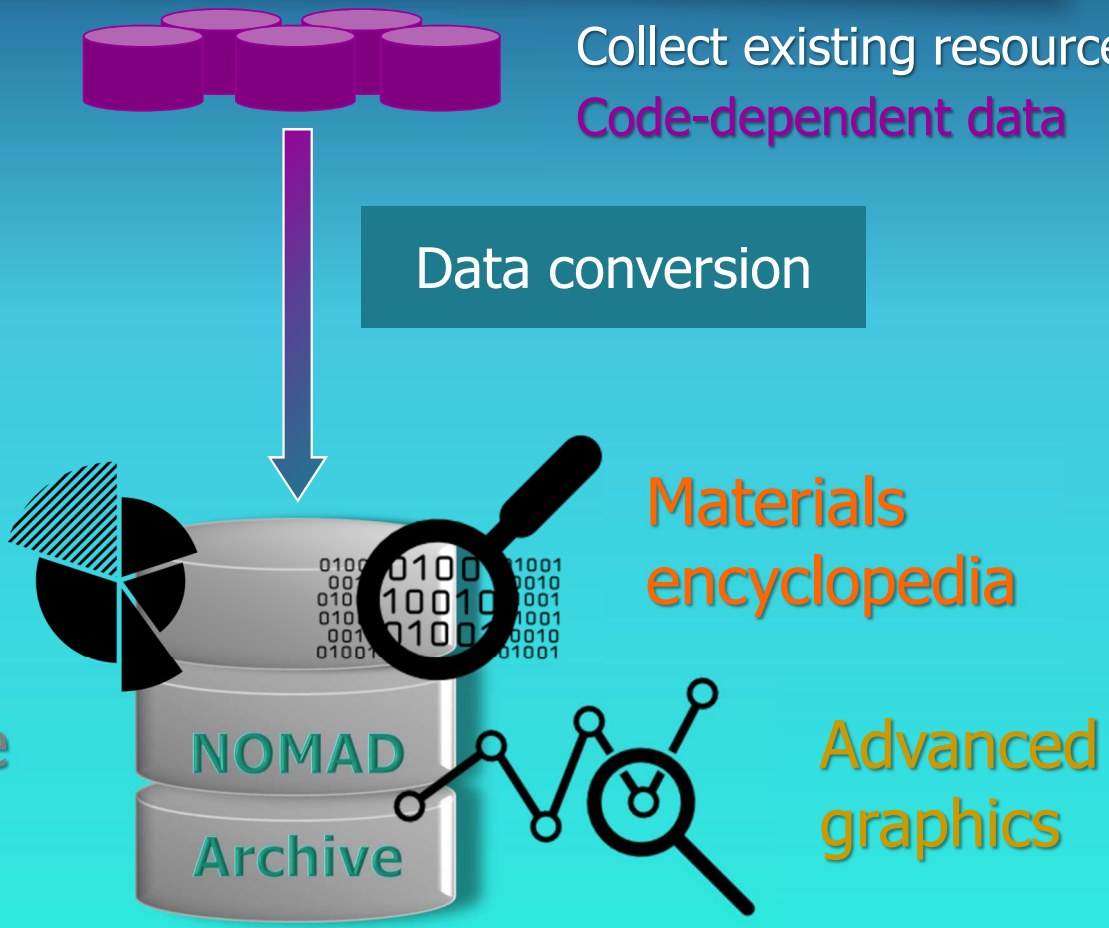
The amount of materials data produced on workstations, compute clusters, and supercomputers is growing exponentially.

Most of it is thrown away ….

The NOMAD Laboratory

Give access to the vast amount of materials data computed worldwide

Get insight from the data

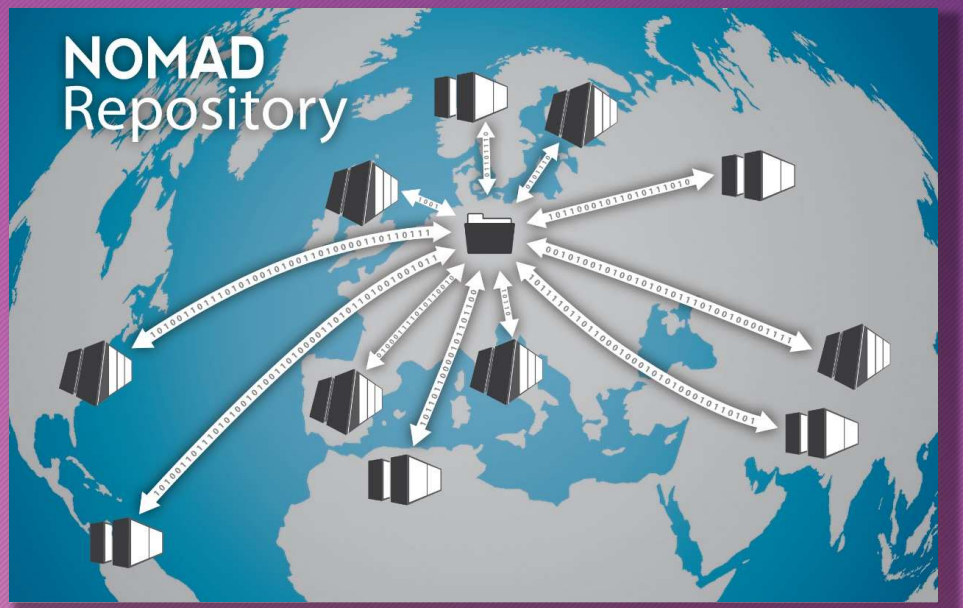


The NOMAD Repository

Data is the raw material of the 21st century

- Host, organize, and share data
  (10 year storage guaranteed)
- In- and output files of all major codes.
- Content (yesterday evening):
  5,045,739 calculations.

https://repository.nomad-coe.eu



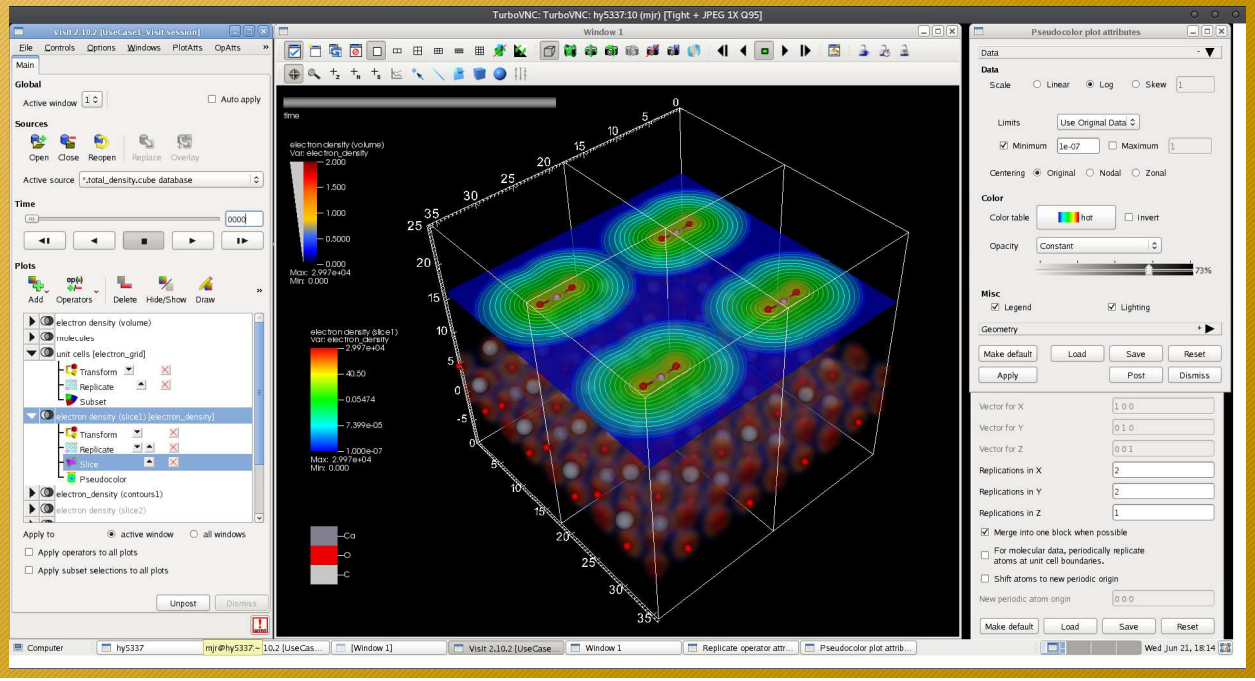
The NOMAD Archive

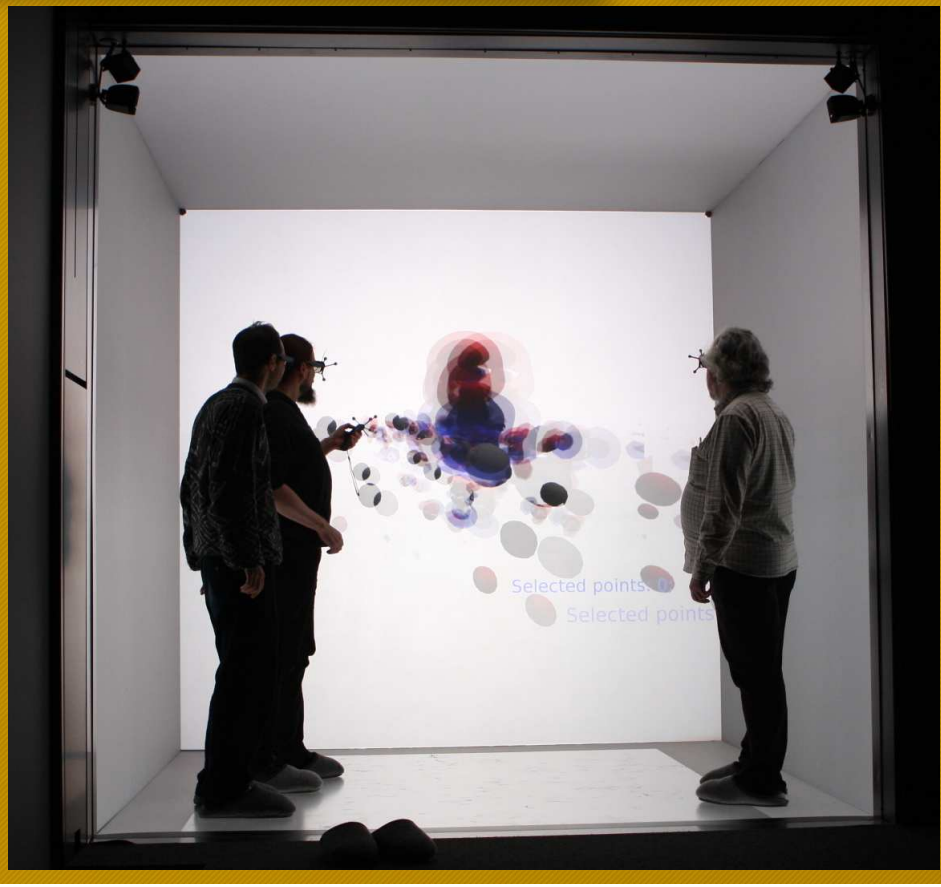
Code-independent representation of the computed properties

- Nomenclature, data representation, and file formats of the input and output files of the community codes are very heterogenous
- Normalization requires definition of **metadata**
  - Generic and code-specific
  - In total >2000 „keys“
- Error bars & trust levels
  - Functionals, force-fields, various approximations …

Advanced visualization

Remote visualization, Virtual Reality, …



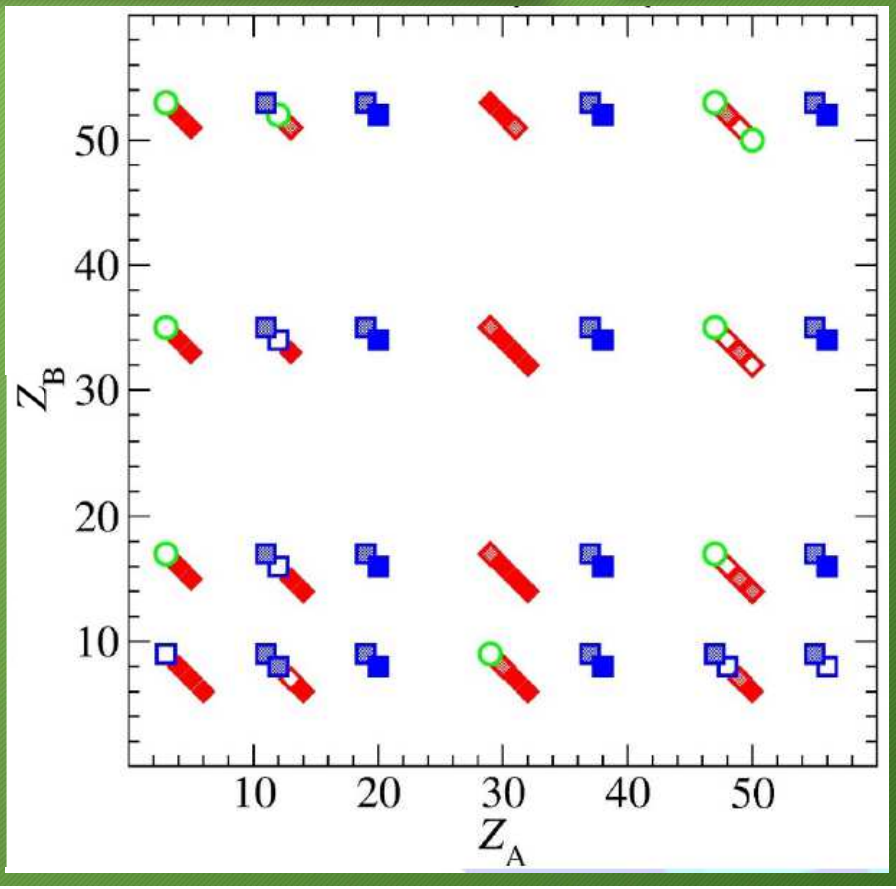


Big-Data Analytics

Example: Phillips - van Fechten problem

Structure map of binary semiconductors

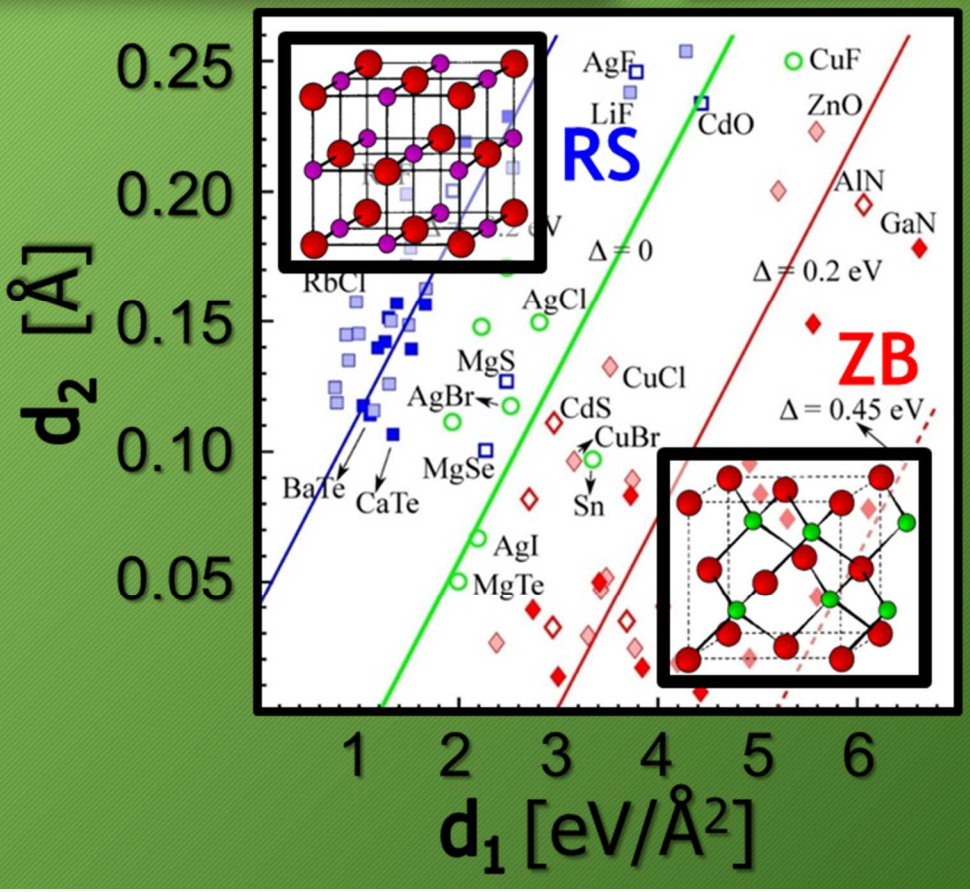
? **Rock Salt** or **Zincblende** ?



Big-Data Analytics

Example: Phillips - van Fechten problem

- Structure map of binary semiconductors, obtained with a compressed-sensing algorithm
- Predictions from free neutral atoms A and B
- Results can be reenacted at https://analytics-toolkit.nomad-coe.eu/

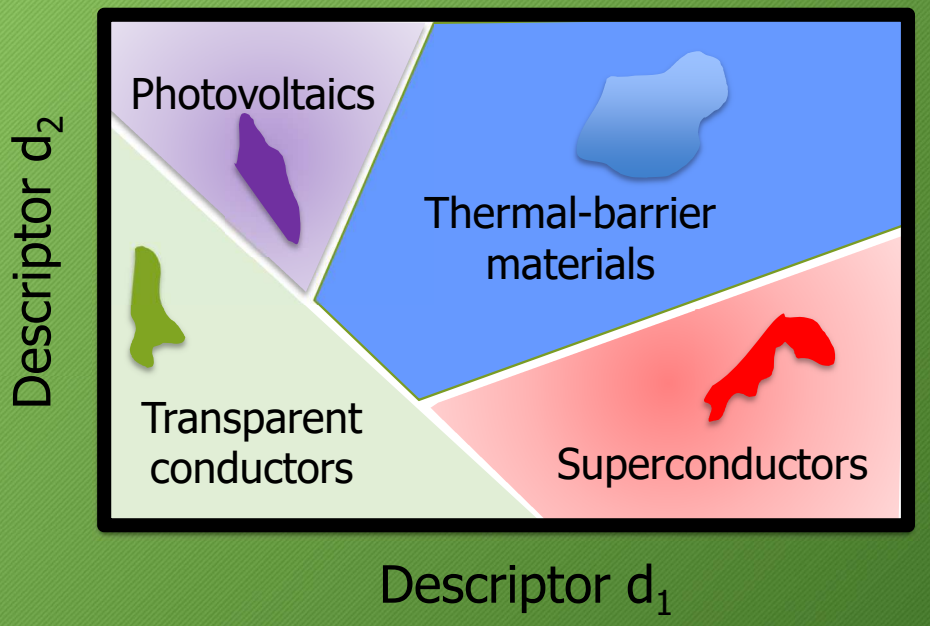


L.M. Ghiringhelli, J. Vybiral, S.V. Levchenko, CD, and M. Scheffler, PRL 114, 105503 (2015).
L.M. Ghiringhelli, et al., New J. Phys. 19, 023017 (2017).

Big-Data Analytics

Identify correlations and structure in the data

Enable scientists and engineers to identify materials for potential use in novel products and to decide which materials should be the focus of future studies.



https://analytics-toolkit.nomad-coe.eu/

The NOMAD Encyclopedia

A **materials-oriented** view on the Archive data
for seeing, comparing, exploring, and understanding.

GUI for **humans**

Rest-API for "**robots**", downloads, …

Current features:

Atomic & electronic structure

Methodology of calculation

Thermal properties

The NOMAD Encyclopedia

Content:

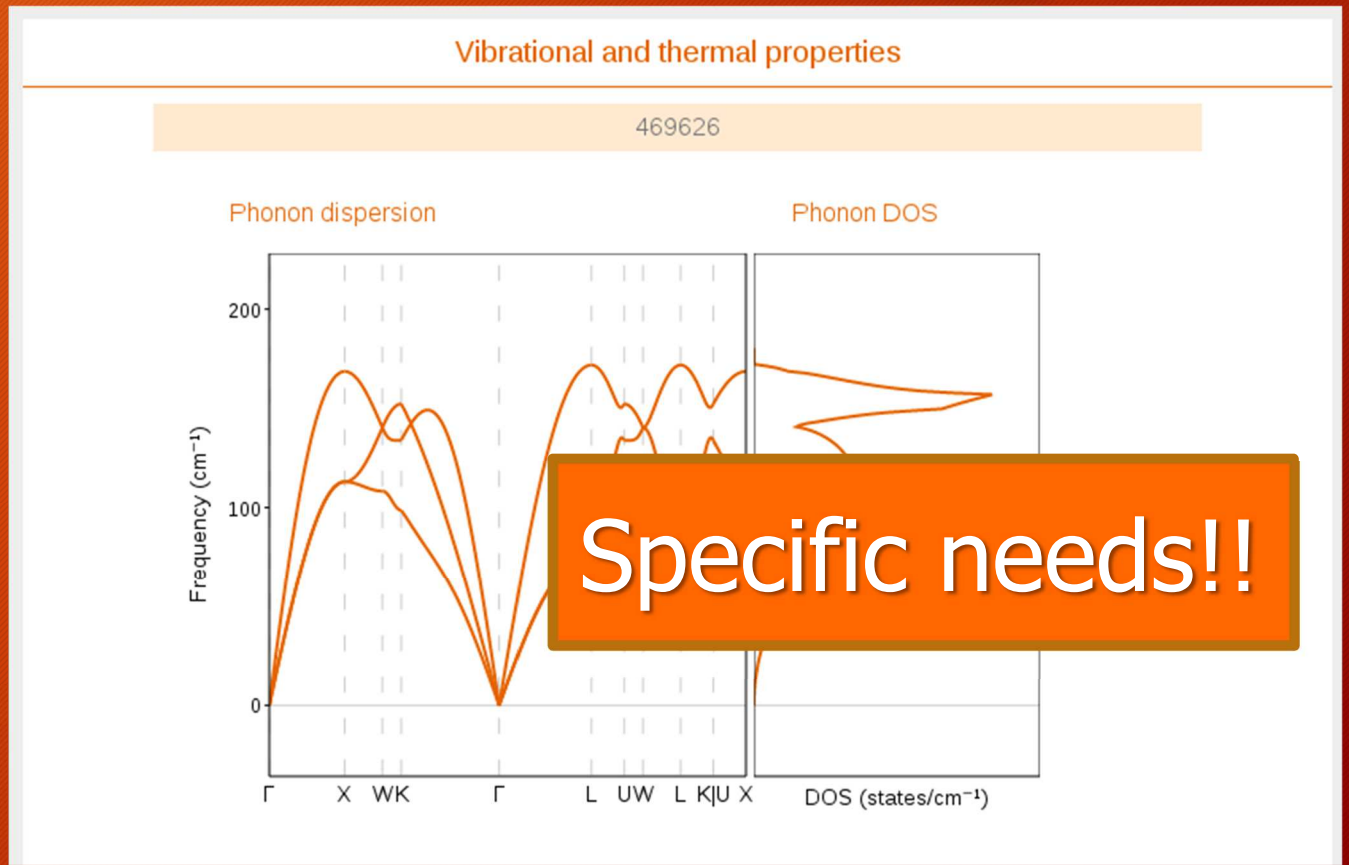
- 229,989 materials
- 2,141,351 calculations

Displayed currently:

- Bulk and 2D systems
- DFT and GW calcs

Highly welcome

- User feedback



The NOMAD Encyclopedia

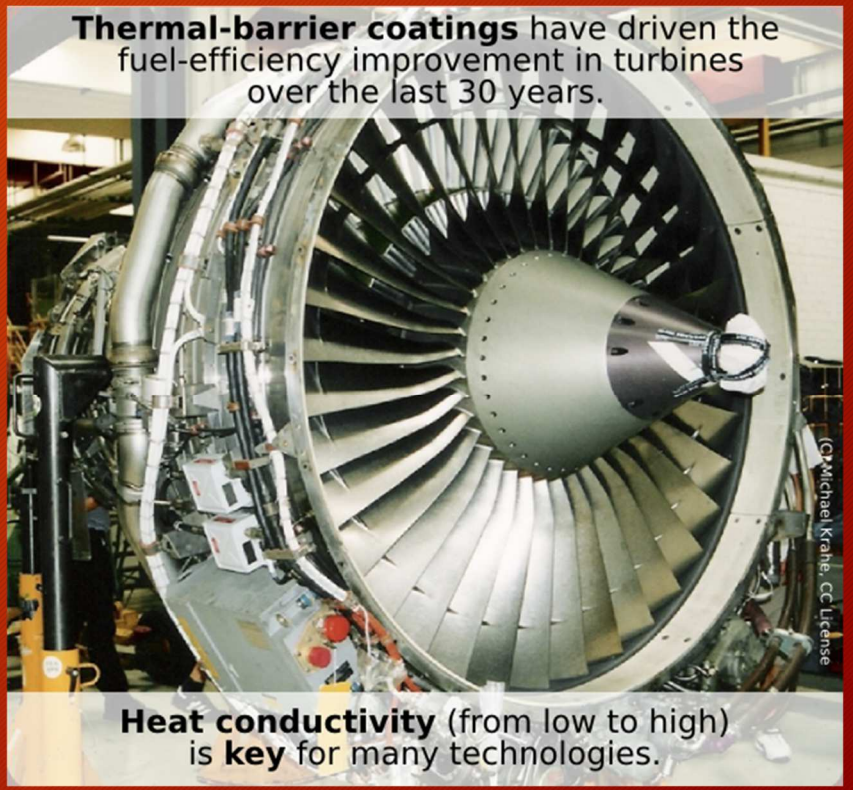
"Specific needs" example

Completely done:

- Phonon based calcs of $C_V(T)$, F(T)

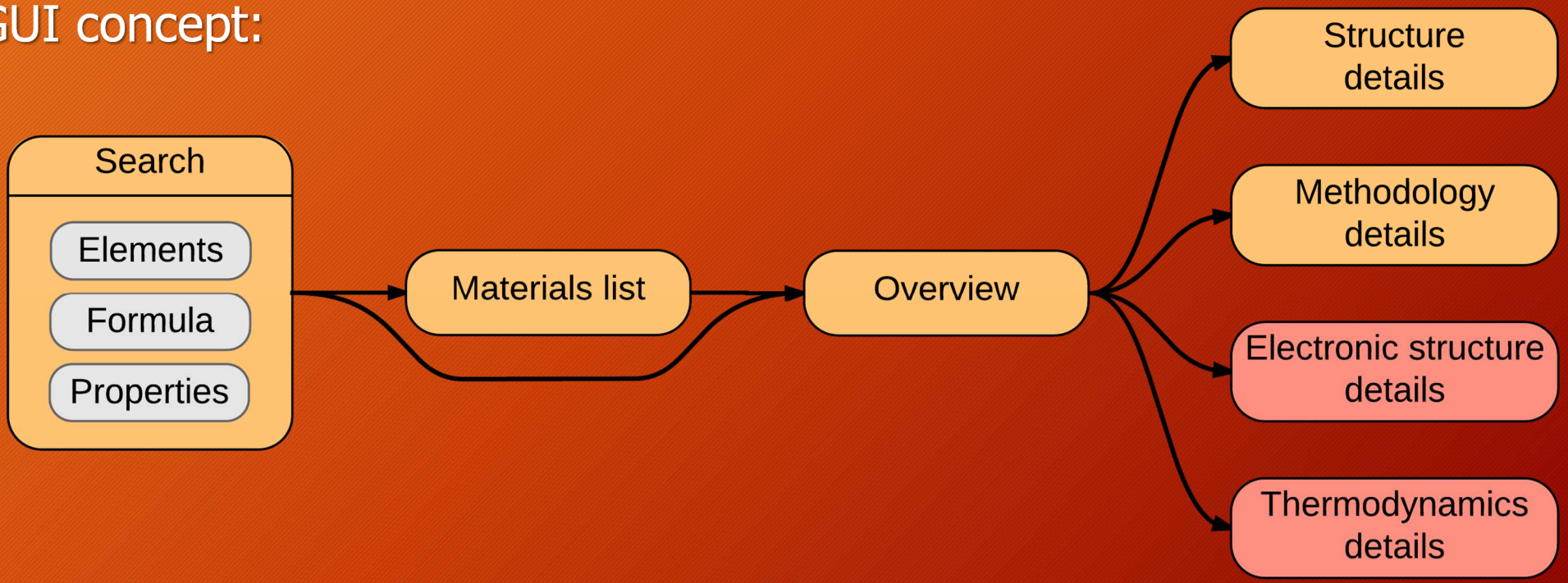
Calcs done:

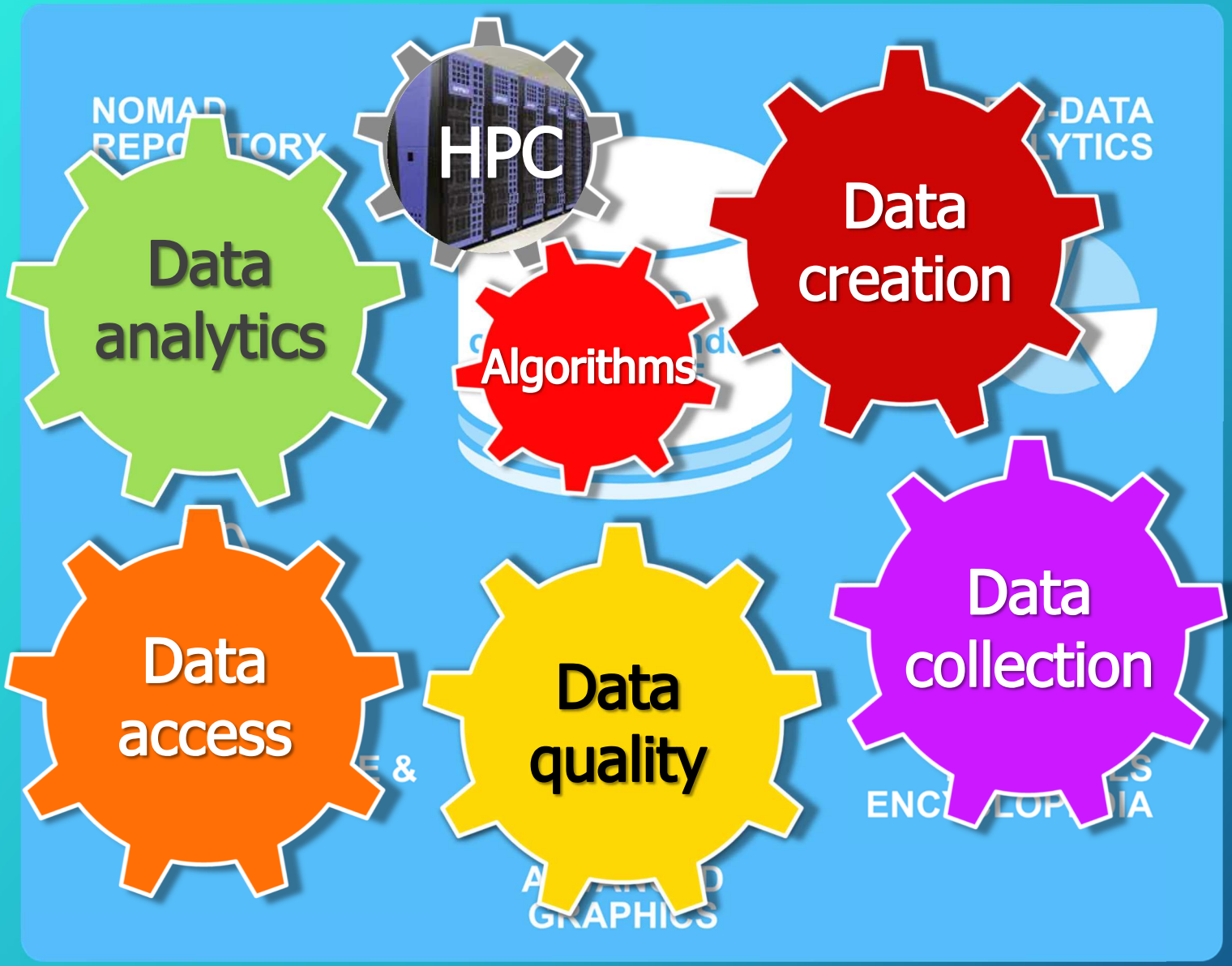
- Quasi-harmonic-approximation based C_V, F, ρ, α_V, K



The NOMAD Encyclopedia

GUI concept:





NOMAD

Thank you !!

Kristian Thygesen
DTU Lyngby

Ciaran Clissman
Pintail Dublin

Arndt Bode
LRZ Munich

Jose Maria Cela
BSC Barcelona

Alessandro De Vita
Kings College London

nomad-coe.eu

Matthias Scheffler
FHI Berlin

NOMAD

Claudia Draxl
HU Berlin

Daan Frenkel
Univ. Cambridge

Angel Rubio
MPSD Hamburg

Risto Nieminen
Aalto Univ. Helsinki

Kimmo Koski
CSC Helsinki

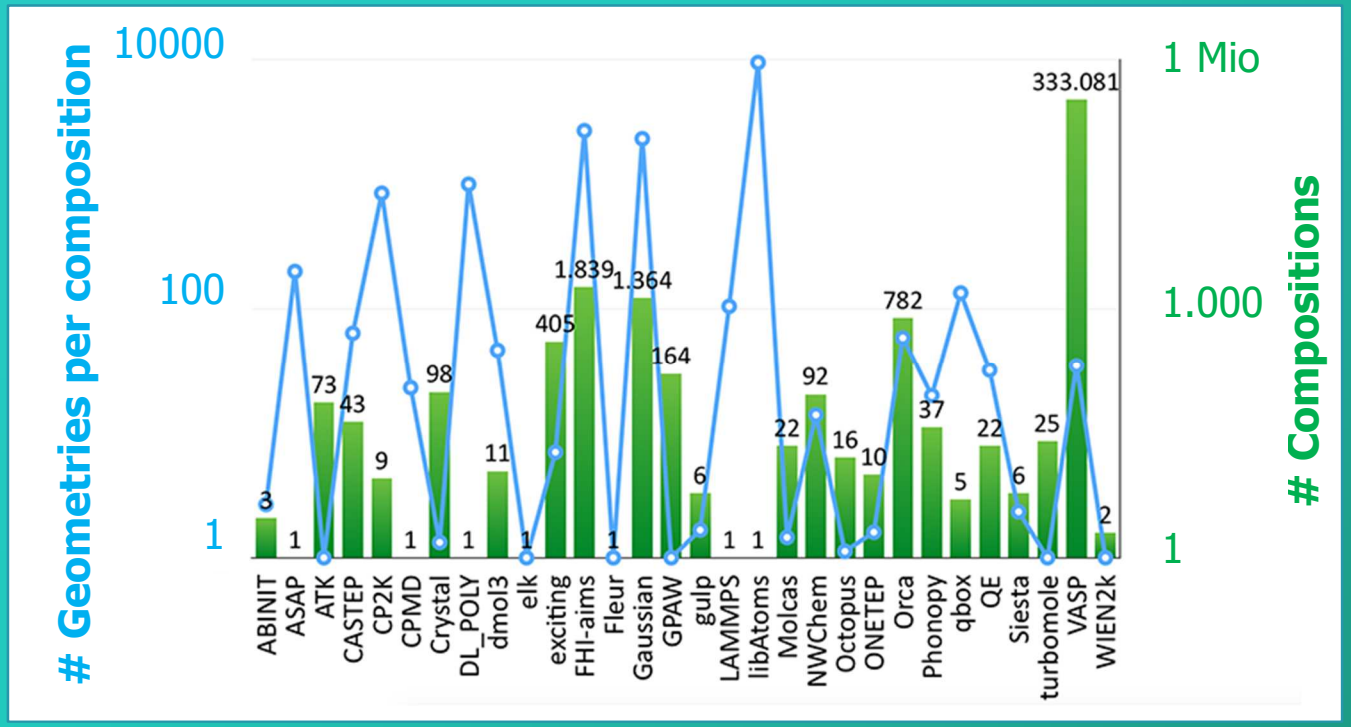
Francesc Illas
Univ. Barcelona

Stefan Heinzel
MPSCD Garching

The NOMAD Archive

https://metainfo.nomad-coe.eu/

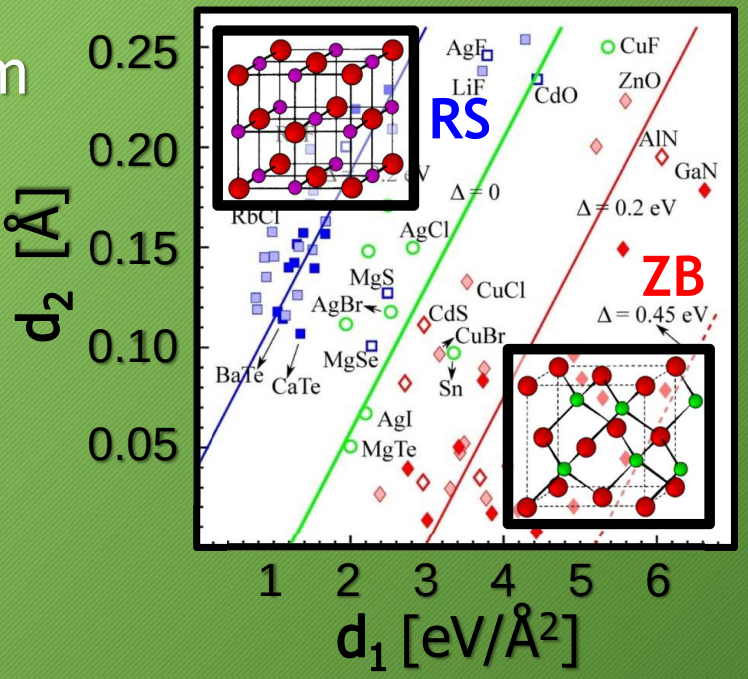
NOMAD supports all important codes



Big-Data Analytics

Example: Phillips - van Fechten problem

- Structure map of binary semiconductors, obtained with a compressed-sensing algorithm
- Predictions from free neutral atoms A and B
- Results can be reenacted at https://analytics-toolkit.nomad-coe.eu/



$$d_1 = \frac{IP(A) - EA(B)}{r_p(A)^2} \qquad d_2 = \frac{|r_s(A) - r_p(B)|}{\exp[rp(A)]}$$

L.M. Ghiringhelli, J. Vybiral, S.V. Levchenko, CD, and M. Scheffler, PRL 114, 105503 (2015).
L.M. Ghiringhelli, et al., New J. Phys. 19, 023017 (2017).