# Data-driven materials research: The NOMAD Laboratory

Claudia Draxl

# Materials data & their structure

| Level | Properties | Methods | Size |
|---|---|---|---|
| I | Atomic positions and nuclear charges, properties of free atoms, symmetry, temperature, pressure | Input: definition of material | 10 kB - 1 MB |
| II | | | 10 MB - 10 TB |
| III | | | 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 | 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 ....

# Thermoelectrics

Direct conversion betw
difference and electric

Enc y





100 MW

Thermo

# Waste-heat recovery

## What makes a good thermoelectric?

Figure of merit ZT with

$$Z = \frac{\sigma S^2}{\kappa_{el}^0 + \kappa_{ph}}$$

S Seebeck coefficient
σ electronic conductivity
κ thermal conductivity

Current values: Z = 0.6 – 1.5

Profitable applications: Z > 2

## Problem:

High electrical conductivity σ and low thermal conductivity κ is required at the same time

# Materials data & their structure

| Level | Properties | Methods | Size |
|---|---|---|---|
| I | Atomic positions and nuclear charges, properties of free atoms, symmetry, temperature, pressure | Input: definition of material *gene* | 10 kB - 1 MB |
| II | Total energy, electron density, potential, wavefunctions, atomic forces, optimized geometry, elastic constants, etc. | Density-functional theory (DFT) and *ab initio* molecular dynamics (MD) | 10 MB - 10 TB |
| III | Excitation energies, dielectric screening, matrix elements of Coulomb interaction, etc. optical spectra, electrical conductivity, phonon spectra, thermal conductivity, etc. | Many-body perturbation theory (MBPT), DF perturbation theory, *ab initio* MD | 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 |

# Novel Materials Discovery

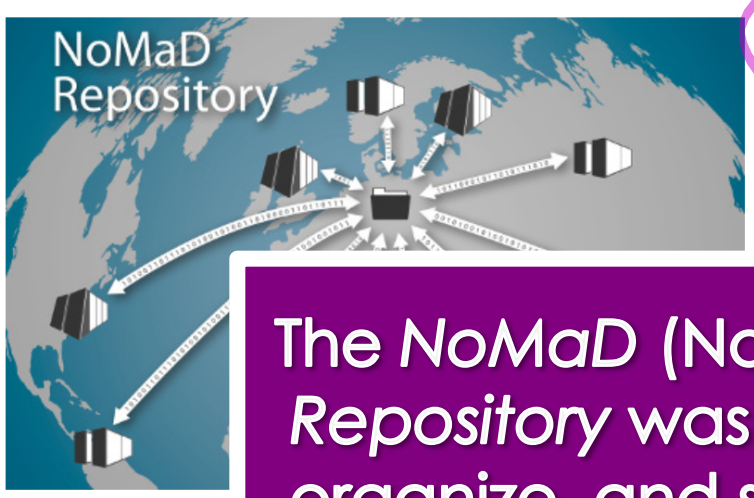
http://nomad-repository.eu

The *NoMaD Repository*

*Insight by sharing*

Home | NoMaD Team | Why sharing? | DOIs | Terms | FAQ | Other repositories | Upload your files | Search and download | Contact us

## Welcome to the NoMaD Repository




**The *NoMaD* (Novel Materials Discovery) *Repository* was established to host, organize, and share materials data.**

***NoMaD*** copes with the increasing demand and requirement of storing scientific data and making them available for longer periods. Rules of good scientific

The ***NoMaD Reposi***
**youtube** to see our

## News

Currently, the NoMaD Repository contains

**3 309 778**

entries.

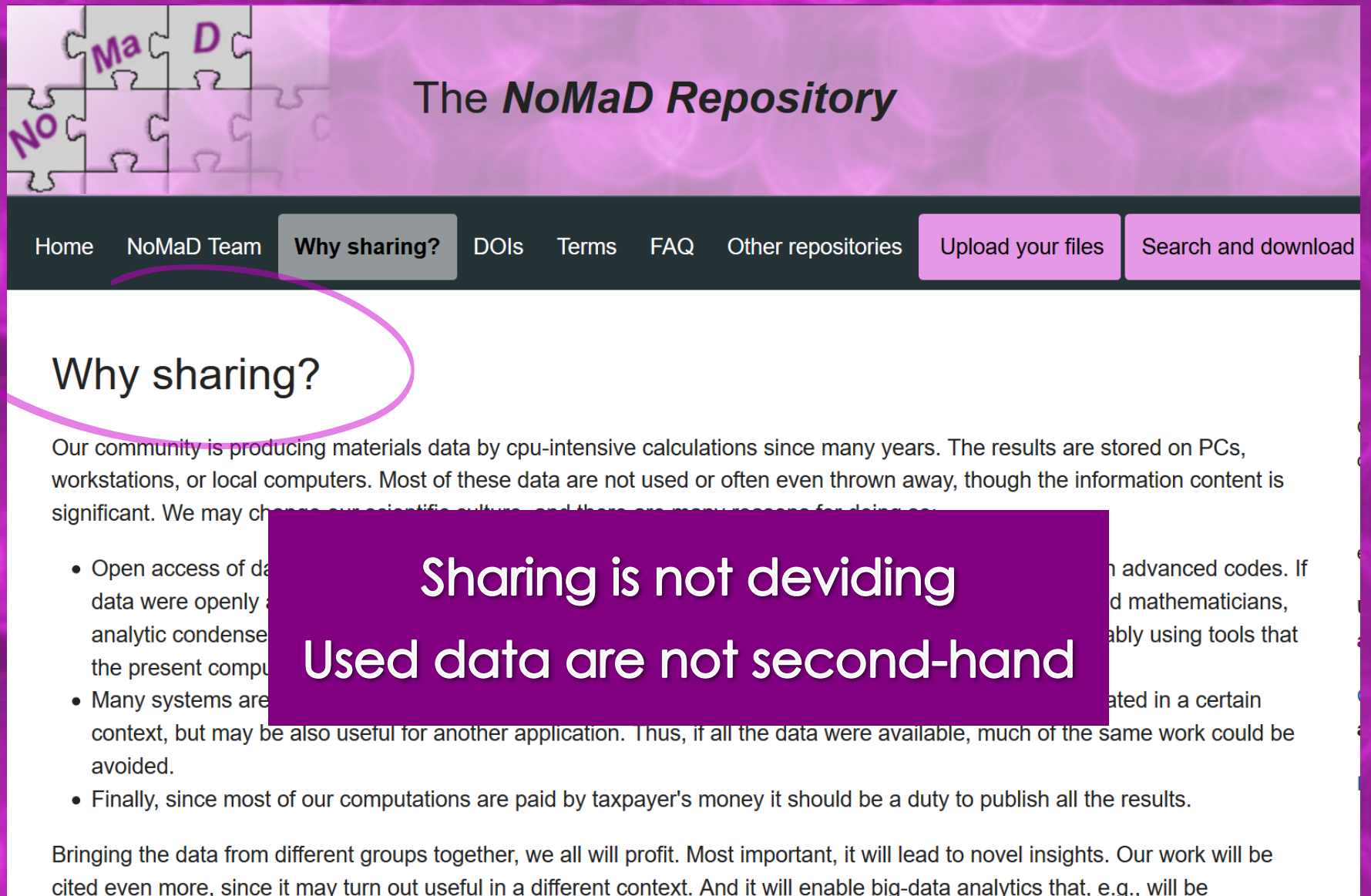
**Upload to *NoMaD* from *MedeA***
**...tion** ... more

...r related **conferences**
**...kshops**.

**...al Support**

The *NoMaD* (Novel Materials Discovers) *Repository* was established to host, organize, and share materials data.

# NoMaD Repository



# Why sharing?

## Avoid doubling of work

Spend human and computational resources for work beyond

Complement the small amount of published information

## Repurposing

One material may be also useful for different applications

Anyone from neighboring research areas can use data

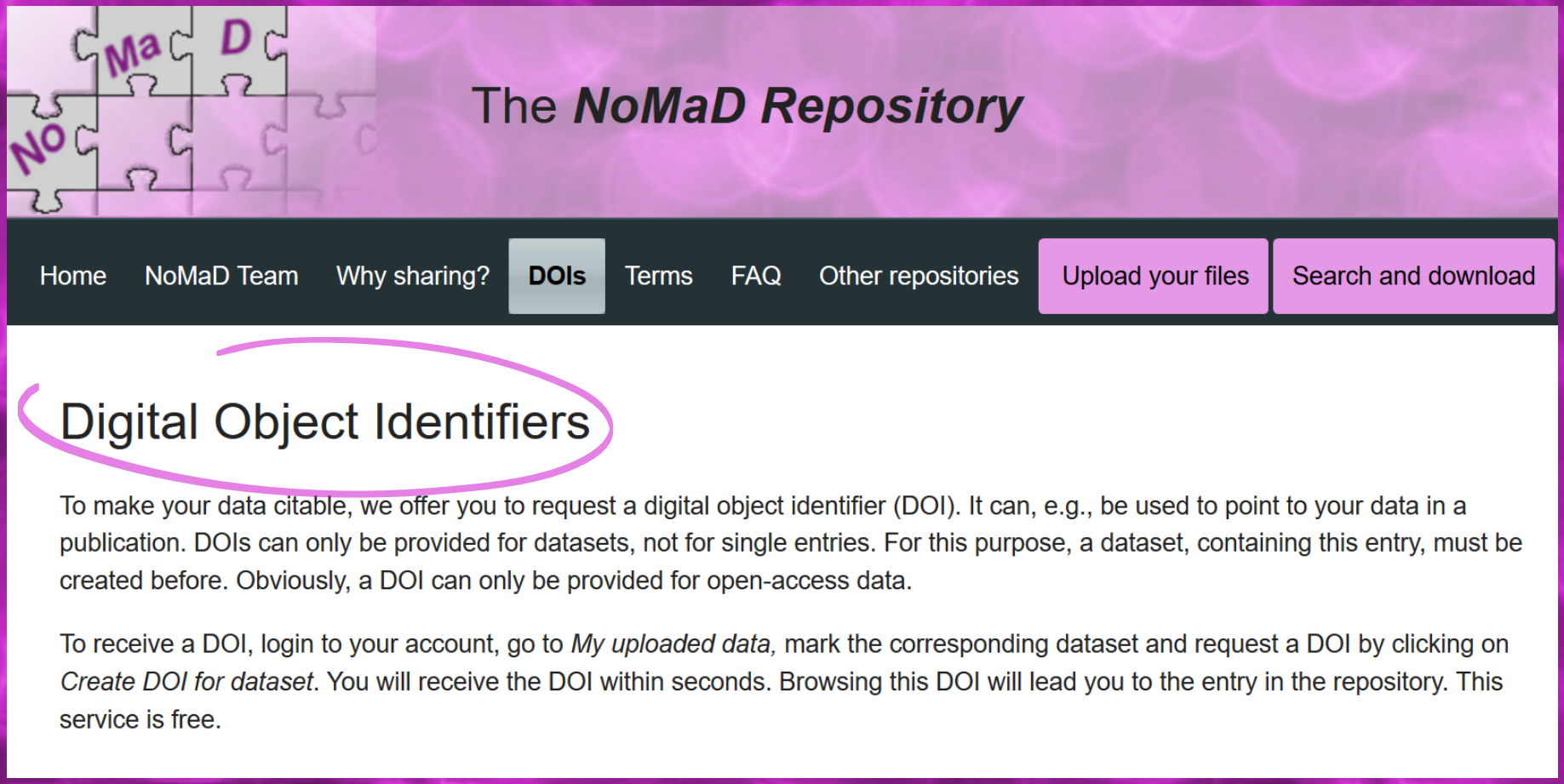
## Good scientific practice

Science foundations require to keep data for 10 years

## Responsibilty

Tax payer's money

# Making data citable




The *NoMaD Repository* is a joint effort by the groups of Matthias Scheffler, FHI Berlin and Claudia Draxl, HU Berlin, and the Computer Center of the Max-Planck Society.

# More aspects of data curation …

- The NoMaD Repository accepts and requests input and output files from all major community codes
  - Currently 20, in a few days 30, in a few months 40
- Only few metadata
  - Uploader, code & version, space group, …
- All data are *valid*
  - Data are produced for a given purpose
  - Errata are possible
- Data are kept for at least 10 years
  - If protected, made open access after 3 years

# What to do with all these data?

Insight by sharing

What can be considered a big-data problem is, actually, a chance - the chance to learn from these data and obtain unprecedented insight.

Besides high-throughput screening, the availability of materials data opens new routes for basic materials science, identifying trends, mechanisms, and anomalies.

**Kristian Thygesen**
DTU Lyngby

**Ciaran Clissman**
Pintail Dublin

**Arndt Bode**
LRZ Munich

**Jose Maria Cela**
BSC Barcelona

**Alessandro De Vita**
Kings College London

**Matthias Scheffler**
FHI Berlin

NOMAD

**Claudia Draxl**
HU Berlin

**Angel Rubio**
MPSD Hamburg

**Risto Nieminen**
Aalto Univ. Helsinki

**Kimmo Koski**
CSC Helsinki

**Francesc Illas**
Univ. Barcelona

**Stefan Heinzel**
MPSCD Garching

**Daan Frenkel**
Univ. Cambridge

# NOMAD CoE

## Aims at filling white spots

https://NOMAD.CoE.eu

NOMAD

# THE NOMAD LABORATORY

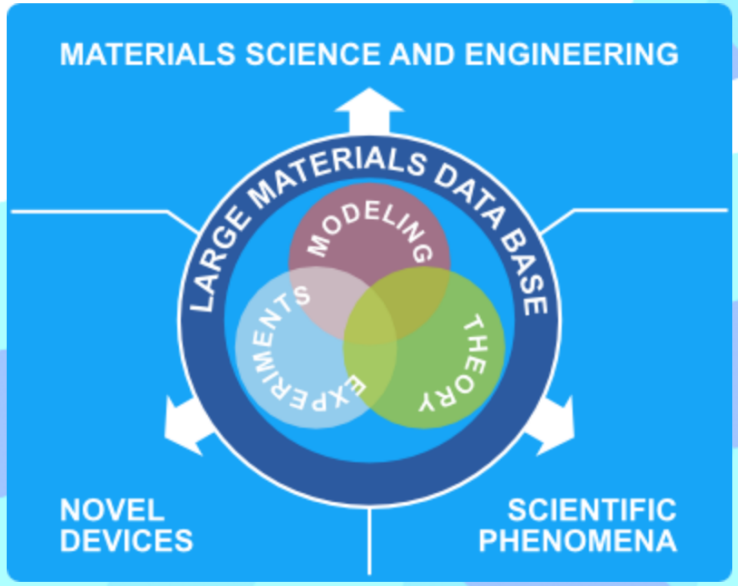
## A EUROPEAN CENTRE OF EXCELLENCE

HOME PROJECT INDUSTRY TEAM CODES NEWS PRESS KIT CONTACT US

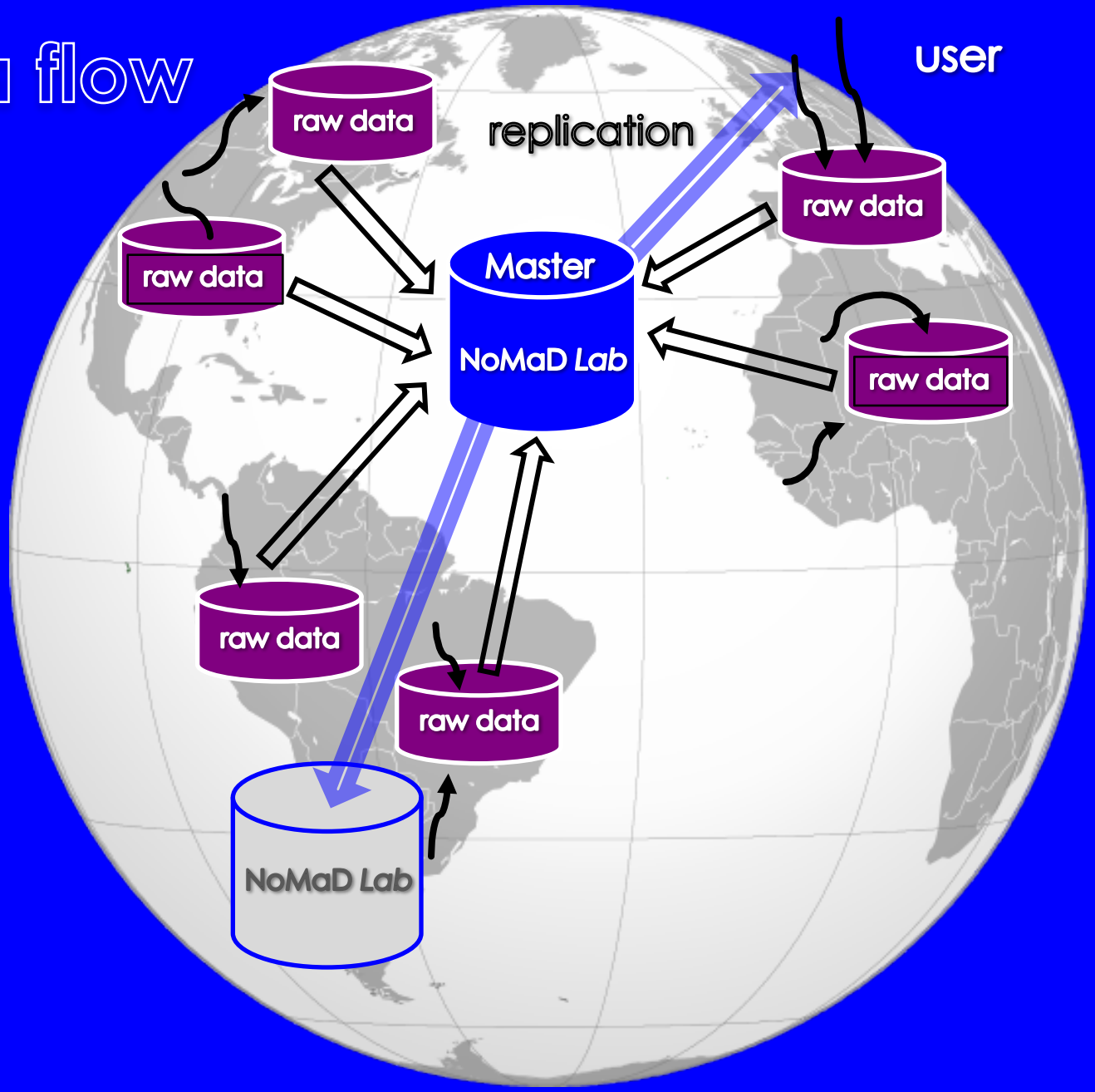
Enter Search...

**The Novel Materials Discovery (NOMAD) Laboratory** develops a *Materials Encyclopedia* and *Big-Data Analytics* and *Advanced Graphics Tools* for materials science and engineering.

Eight complementary computational materials science groups and four high-performance computing centers form the synergetic core of this Centre of Excellence.

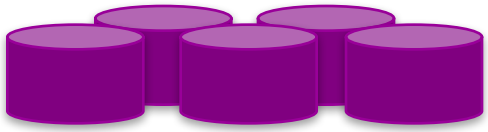


# Data flow


user
raw data
replication
raw data
raw data
Master
NoMaD Lab
raw data
raw data
raw data
NoMaD Lab

# NOMAD Laboratory

Existing resources
Code-dependent data

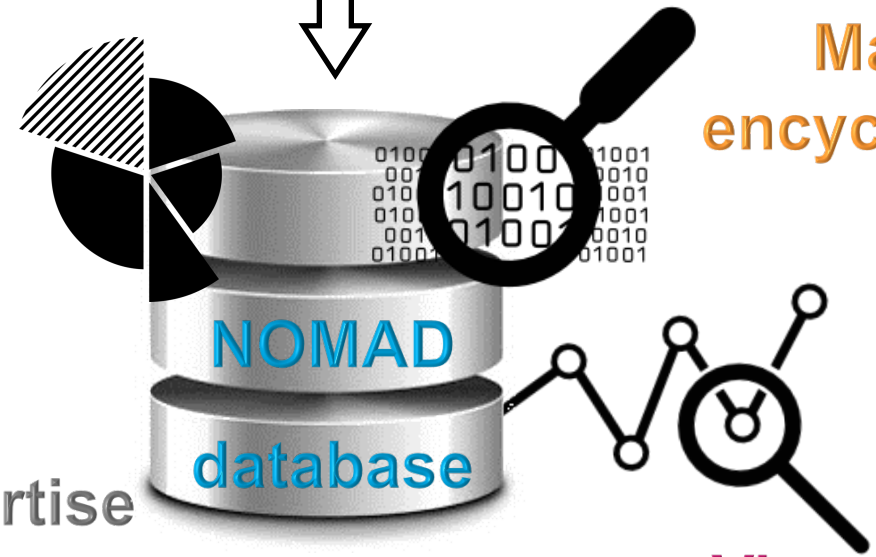


Data conversion

Give access to the vast *amount* of materials data computed worldwide

Big-data analytics

Materials encyclopedia




HPC expertise & hardware

Visualization

# Data conversion

## How to make data comparable?

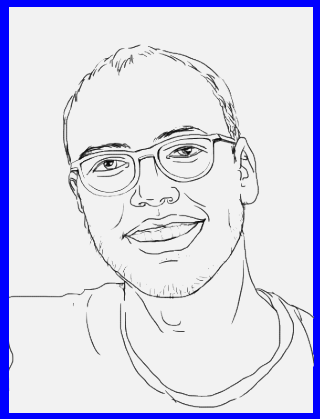
- NOMAD supports ~40 different computer codes
- Common representation for various quantities
  - Pseudopotentials vs all-electron methods, …
- Evaluate error bars
  - Different functionals, force fields, …

## Metadata

- Hierarchical metadata schema
- Generic and code-specific
  - https://nomad-coe.eu/index.php?page=nomad-meta-info

# What about the data quality?

## Validation



Sven Lubeck

Andris Gulans

# Delta factors

RESEARCH ARTICLE

DFT METHODS

# Reproducibility in density functional theory calculations of solids

Kurt Lejaeghere,[1*] Gustav Bihlmayer,[2] Torbjörn Björkman,[3,4] Peter Blaha,[5] Stefan Blügel,[2] Volker Blum,[6] Damien Caliste,[7,8] Ivano E. Castelli,[9] Stewart J. Clark,[10] Andrea Dal Corso,[11] Stefano de Gironcoli,[11] Thierry Deutsch,[7,8] John Kay Dewhurst,[12] Igor Di Marco,[13] Claudia Draxl,[14,15] Marcin Dułak,[16] Olle Eriksson,[13] José A. Flores-Livas,[12] Kevin F. Garrity,[17] Luigi Genovese,[7,8] Paolo Giannozzi,[18] Matteo Giantomassi,[19] Stefan Goedecker,[20] Xavier Gonze,[19] Oscar Grånäs,[13,21] E. K. U. Gross,[12] Andris Gulans,[14,15] François Gygi,[22] D. R. Hamann,[23,24] Phil J. Hasnip,[25] N. A. W. Holzwarth,[26] Diana Iuşan,[13] Dominik B. Jochym,[27] François Jollet,[28] Daniel Jones,[29] Georg Kresse,[30] Klaus Koepernik,[31,32] Emine Küçükbenli,[9,11] Yaroslav O. Kvashnin,[13] Inka L. M. Locht,[13,33] Sven Lubeck,[14] Martijn Marsman,[30] Nicola Marzari,[9] Ulrike Nitzsche,[31] Lars Nordström,[13] Taisuke Ozaki,[34] Lorenzo Paulatto,[35] Chris J. Pickard,[36] Ward Poelmans,[1,37]

# Delta factors

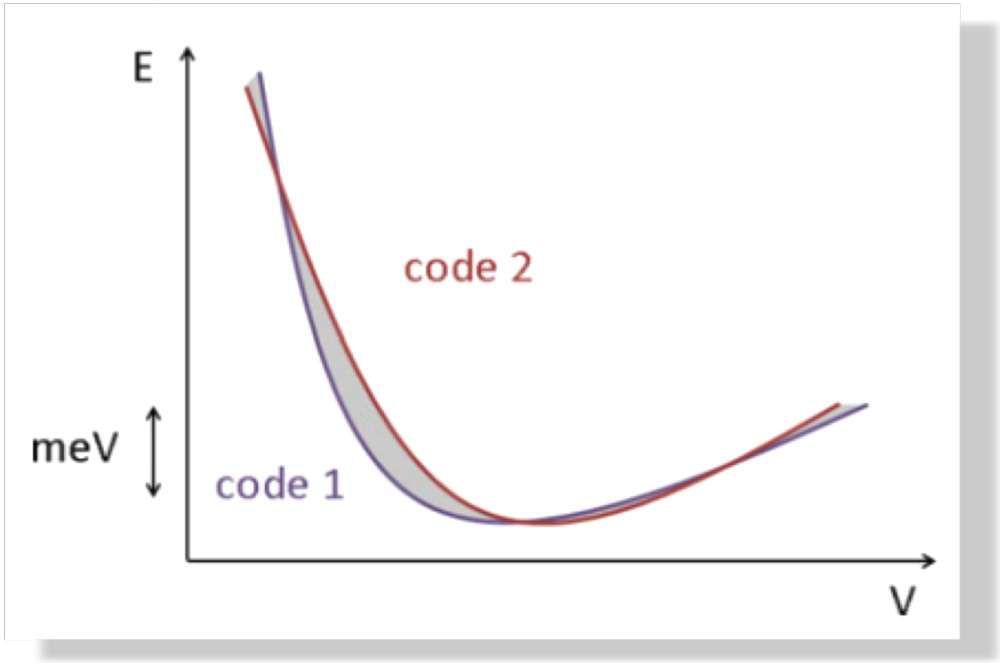
Compute E(V) using PBE

Fit to the Birch-Murnaghan equation of state

Compare with other codes / method

## Quality factor Δ

$$\Delta = \left\langle \sqrt{\frac{\int \Delta E^2(V)\mathrm{d}V}{\Delta V}} \right\rangle$$




RESEARCH ARTICLE SUMMARY

DFT METHODS

**Reproducibility in density functional theory calculations of solids**

K. Lejaeghere et al.,
Science **351**, aad3000 (2016).

# Delta factors

https://molmod.ugent.be/deltacodesdft

| Code | Version | Basis | Electron treatment | Δ-value | Authors |
|---|---|---|---|---|---|
| Exciting | development version | LAPW+xlo | all-electron | 0 meV/atom | Exciting [10,16] |
| FHI-aims | 081213 | tier2 numerical orbitals | all-electron (relativistic atomic_zora scalar) | 0.1 meV/atom | ASE [2,16] |
| WIEN2k | 13.1 | LAPW/APW+lo | all-electron | 0.2 meV/atom | S. Cottenier [16] |
| FHI-aims | 081213 | tier2 numerical orbitals | all-electron (relativistic zora scalar 1e-12) | 0.3 meV/atom | ASE [2] |
| Quantum ESPRESSO | 5.1 | plane waves | SSSP Accuracy (mixed NC/US/PAW potential library) | 0.3 meV/atom | QuantumESPRESSO [12,16] |

# Delta factors

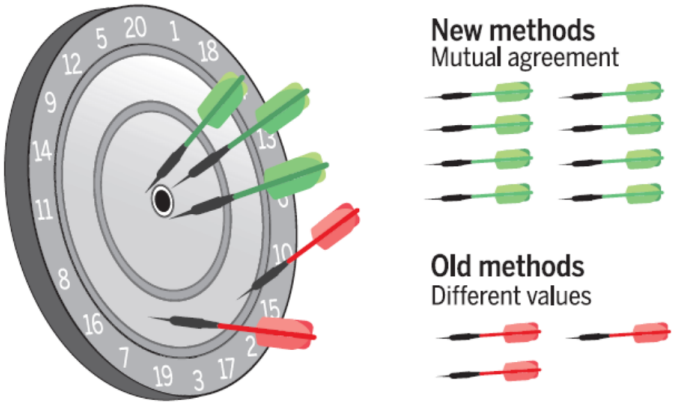
https://molmod.ugent.be/deltacodesdft

| Code | Version | Basis | Electron treatment | Δ-value | Authors |
|---|---|---|---|---|---|
| Exciting | development version | LAPW+xlo | all-electron | 0 meV/atom | Exciting [10,16] |
| FHI-aims | 081213 | tier2 numerical orbitals | all-electron (relativistic atomic_zora scalar) | 0.1 meV/atom | ASE [2,16] |
| WIEN2k | 13.1 | LAPW/APW+lo | all-electron | 0.2 meV/atom | S. Cottenier [16] |
| FHI-aims | 081213 | tier2 numerical orbitals | all-electron (relativistic zora scalar 1e-12) | 0.3 meV/atom | ASE [2] |
| Quantum ESPRESSO | 5.1 | plane waves | SSSP Accuracy (mixed NC/US/PAW potential library) | 0.3 meV/atom | QuantumESPRESSO [12,16] |

# Delta factors

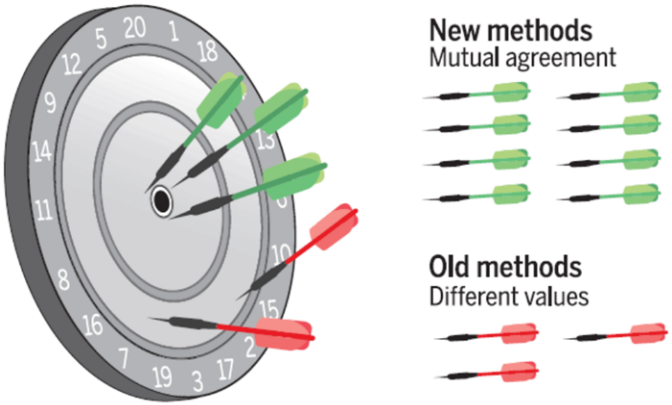
K. Lejaeghere et al.,
Science **351**, aad3000 (2016).

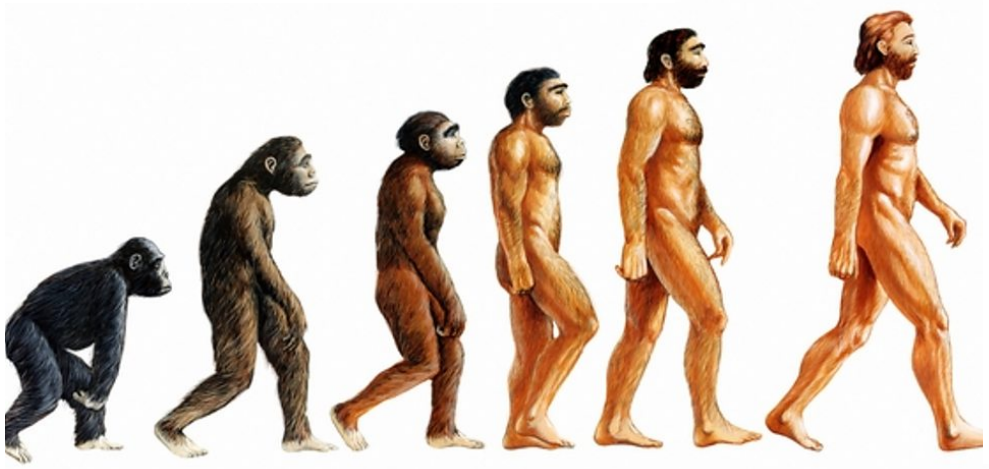
$O_2$




| | exciting | FHI-AIMS/tier2 | FLEUR | FPLO/T+F+s | RSPt | WIEN2k/acc | VASP2011/VASP |
|---|---|---|---|---|---|---|---|
| exciting | | 0.1 | 0.4 | 1.8 | 0.7 | 0.6 | 7.7 |
| FHI-AIMS/tier2 | 0.1 | | 0.3 | 1.7 | 0.6 | 0.5 | 7.6 |
| FLEUR | 0.4 | 0.3 | | 1.4 | 0.3 | 0.3 | 7.2 |
| FPLO/T+F+s | 1.8 | 1.7 | 1.4 | | 1.1 | 1.2 | 5.8 |
| RSPt | 0.7 | 0.6 | 0.3 | 1.1 | | 0.1 | 6.9 |
| WIEN2k/acc | 0.6 | 0.5 | 0.3 | 1.2 | 0.1 | | 7.1 |
| VASP2011/VASP | 7.7 | 7.6 | 7.2 | 5.8 | 6.9 | 7.1 | |

# Delta factors



| | year | $\langle\Delta\rangle$ vs AE |
|---|---|---|
| JTH01/ABINIT | 2013 | 1.1 |
| JTH02/ABINIT | 2014 | 0.6 |
| Vdb/CASTEP | 1998 | 6.5 |
| OTFG7/CASTEP | 2013 | 2.6 |
| OTFG9/CASTEP | 2015 | 0.7 |
| GPAW06/GPAW | 2010 | 3.6 |
| GPAW09/GPAW | 2012 | 1.6 |
| PSlib031/QE | 2013 | 1.7 |
| PSlib100/QE | 2013 | 1.0 |
| VASP2007/VASP | 2007 | 2.0 |
| VASP2012/VASP | 2012 | 0.8 |
| VASPGW2015/VASP | 2015 | 0.6 |

Slide from S. Cottenier

K. Lejaeghere et al., Science **351**, aad3000 (2016).

# This is all great

… but just the beginning

What about other systems

surfaces, defects, molecules, …

What about other quantities

band gaps, barriers, spectra …

# Can we reach ultimate precision?

## Total energies of atoms compared to MADNESS

Andris Gulans

exciting

## Same for molecules

## Yes we can!

| | Ha | μHa |
|---|---|---|
| | $E^{\mathrm{LAPW+lo}}$ | $\Delta E^{\mathrm{MRA}}$ |
| H | −0.4787107 | 0.3 |
| He | −2.8344552 | 0.2 |
| Li | −7.3432843 | 1.6 |
| Be | −14.4464735 | 0.6 |
| B | −24.3548568 | 1.3 |
| C | −37.4685404 | 0.9 |
| N | −54.1343867 | 0.7 |
| O | −74.5286993 | 0.2 |
| F | −99.1118530 | −0.3 |
| Ne | −128.2299171 | 0.3 |
| Na | −161.4436320 | 2.1 |
| Mg | −199.1352882 | 0.7 |
| Al | −241.3178300 | 2.3 |
| Si | −288.2171655 | 1.6 |
| P | −340.0000526 | 1.8 |
| S | −396.7390648 | 0.7 |
| Cl | −458.6643433 | 0.1 |
| Ar | −525.9397933 | 0.8 |

# NOMAD Encyclopedia

The Materials Encyclopedia transforms the calculation-oriented information in the *Archive* (Repository) to materials-oriented knowledge and provides a user-friendly public access point to it.

# NOMAD Encyclopedia

## Comprehensive characterization of materials

### Extensive information on a specific material

Structural features

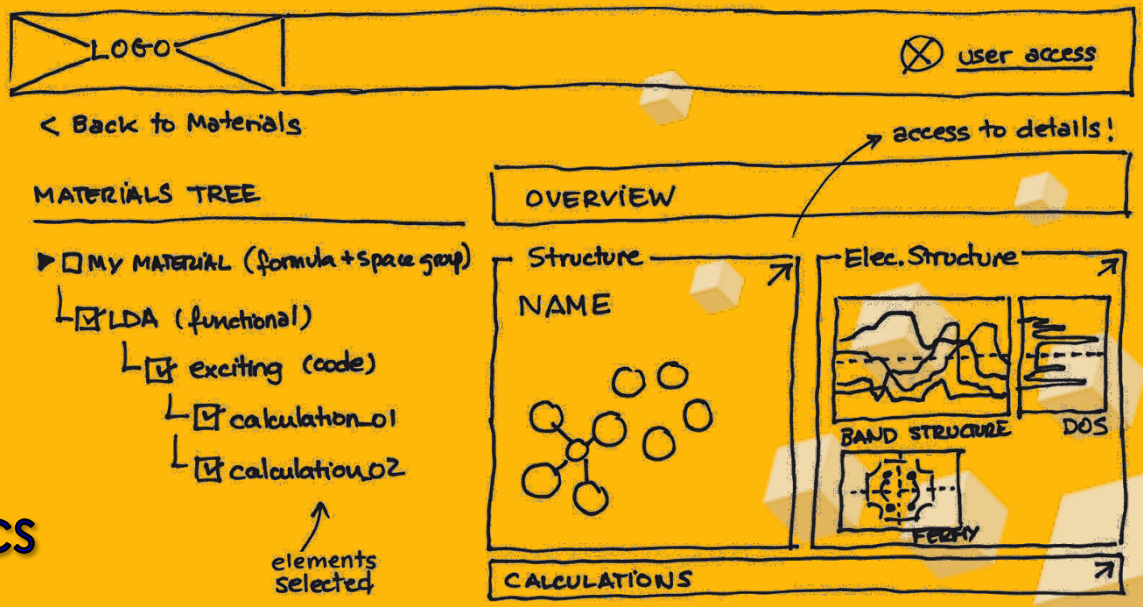
Mechanical behavior

Thermal properties

Electronic structure

Transport characteristics

Response to light and other excitations




Overview & statistics

Methodology & confidence level(s)

Individual calculations & links

# NOMAD Encyclopedia

## Simplest solution: show data

NOMAD Encyclopedia

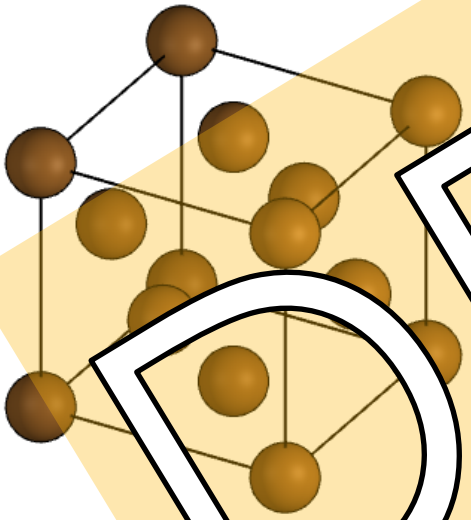
NOMAD user

Results > Overview > **Structure Details**

Copper - Cu - fcc - space group 225

DEMO

Structure



**System type**: bulk
**Structure type**: fcc
**Lattice**: Cubic
**Space group**: 225
**Point group**: m-3m

**Wyckoff positions**
Cu: a

Calculati

Copper
GGA
VA (46)
LDA

Statistics

**Lattice constants** [min, max]:
a = 3.61 Å [3.51, 3.64]
b = 3.61 Å [3.51, 3.64]
c = 3.61 Å [3.51, 3.64]
α = 90 °
β = 90 °
γ = 90 °

**Cell volume** [min, max]:
11.73 Å$^3$ [10.79, 12.05]

**Density** [min, max]:
0.34 Å$^{-3}$ [0.33, 0.37]

ID 2

a = 3.61 Å
b = 3.61 Å
c = 3.61 Å
α = 90 °
β = 90 °
γ = 90 °

**Cell volume**: 11.81 Å$^3$

**Density**: 0.34 Å$^{-3}$

**Energy decomposition**

Statistics

# NOMAD Encyclopedia

## Search for materials that exhibit desired features

Example: Find all materials …

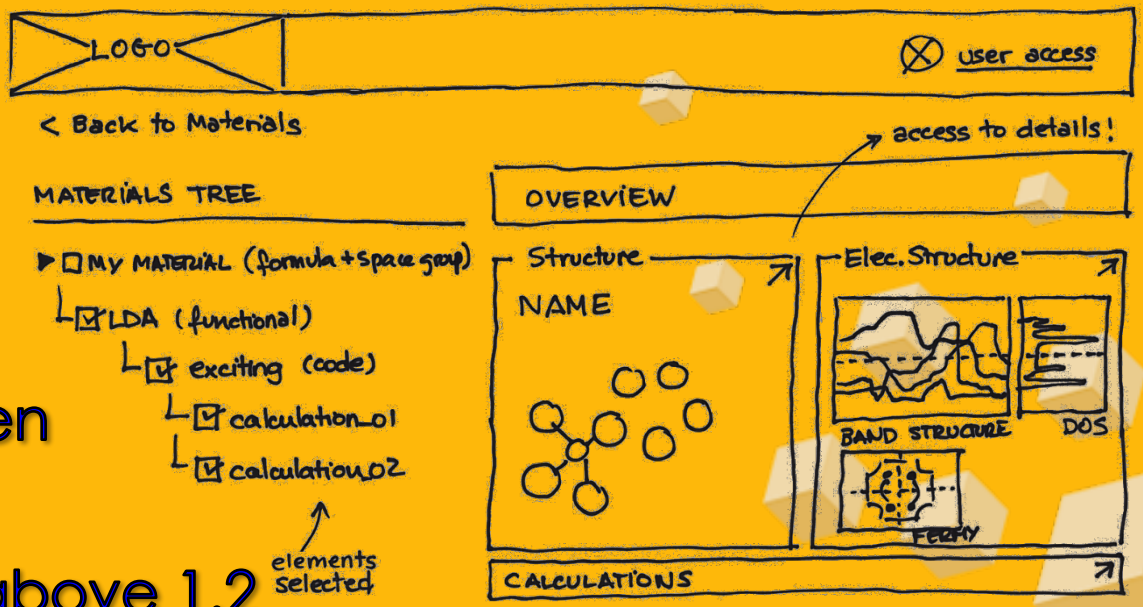
containing C or Si but not Cd or As

ductile below room temperature

with band gap between 1.0 and 1.3 eV

with effective masses above 1.2

without inversion symmetry




Not all information directly available in the Archive

# Data Analytics

Identifying correlations and structure in big data of materials will enable scientists and engineers to decide which materials are useful for specific applications or which new materials should be the focus of future studies.

# Summary and outlook

The NoMaD Repository serves the purpose of organizing and sharing materials data

The NOMAD Laboratory aims at giving access to and getting insight into the vast *amount* of materials data computed worldwide

- Encyclopedia available online May 2017
- First data-analytics tools online already

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

# Thank you!