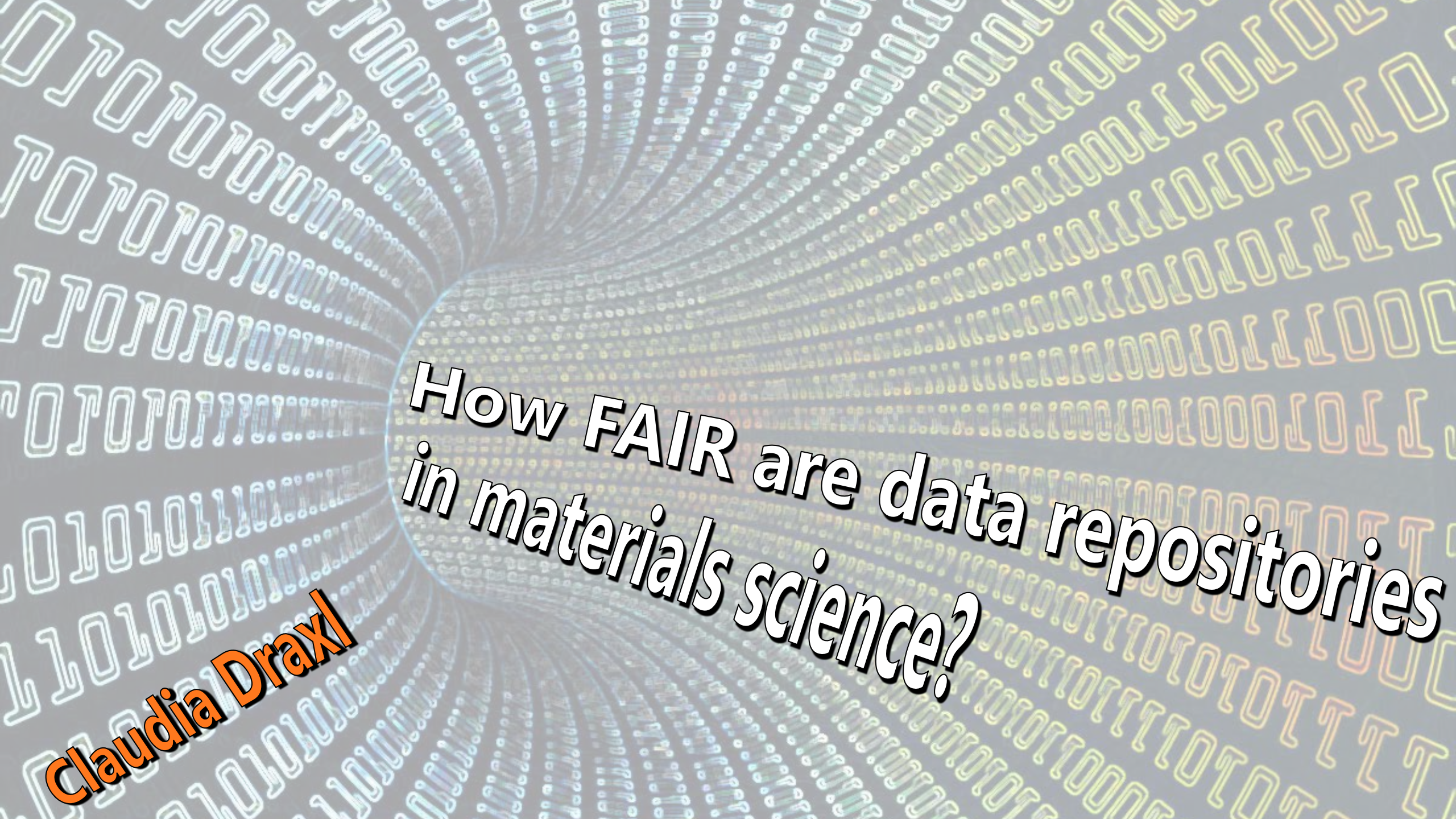
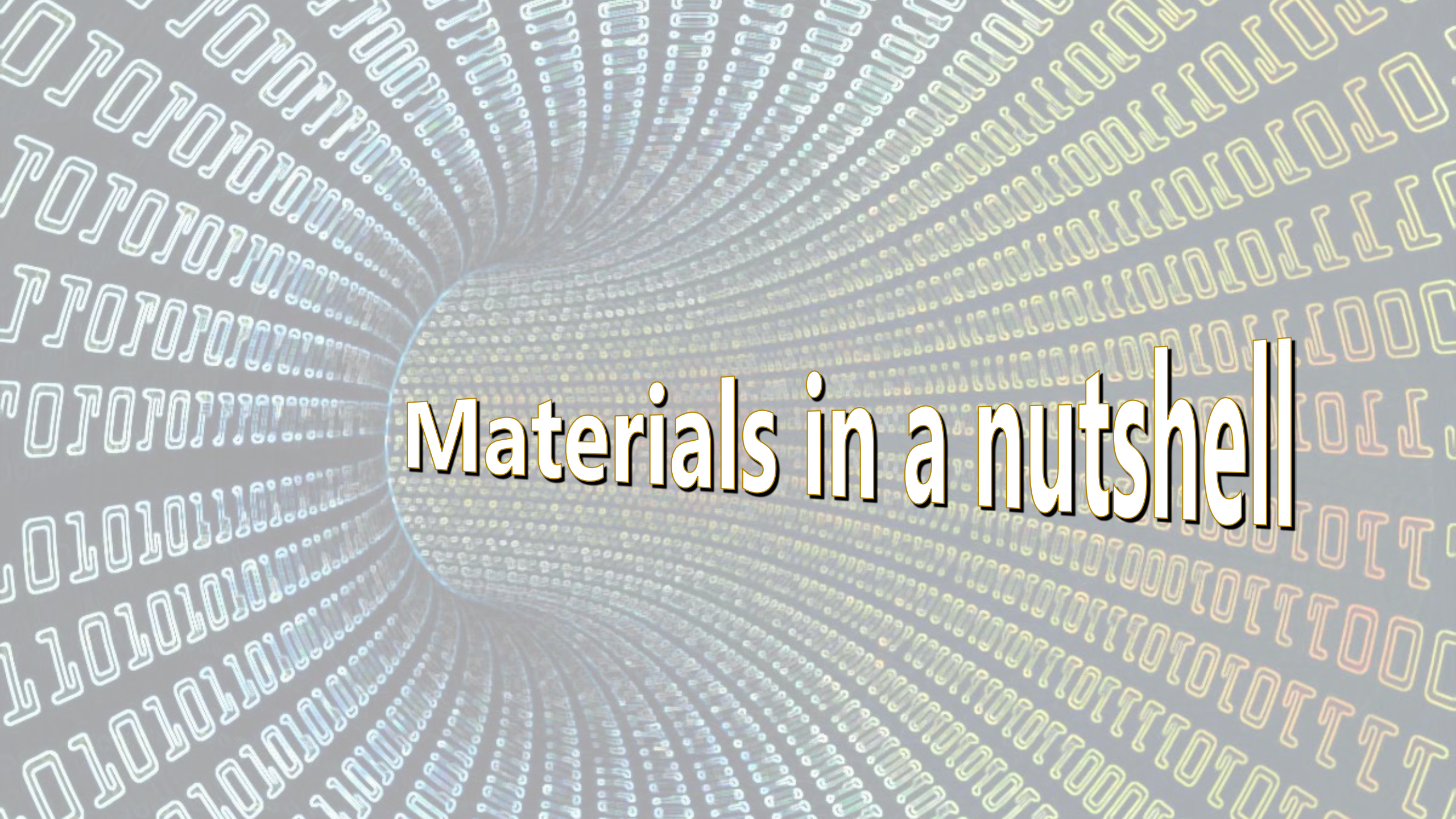

How FAIR are data repositories in materials science?
Claudia Draxl


Materials in a nutshell

It is all about materials ...

Every new product requires novel or improved materials

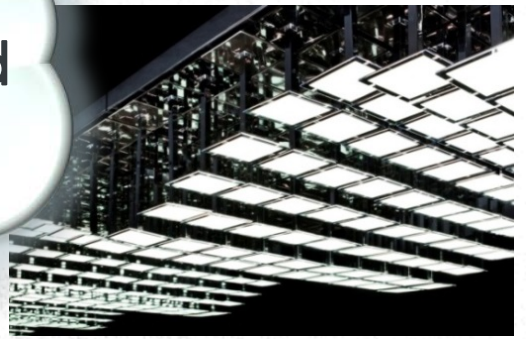
Flexibel, ultra-thin
Nontoxic, sustainable

Light-weight & ductile
Heat-resistant

Biocompatible

Hard & tough

High quantum yield
Energy-saving
Desired color

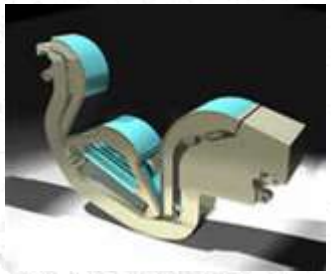


Example: fusion reactors

Tungsten?

Highest melting point: 3683K

Divertor

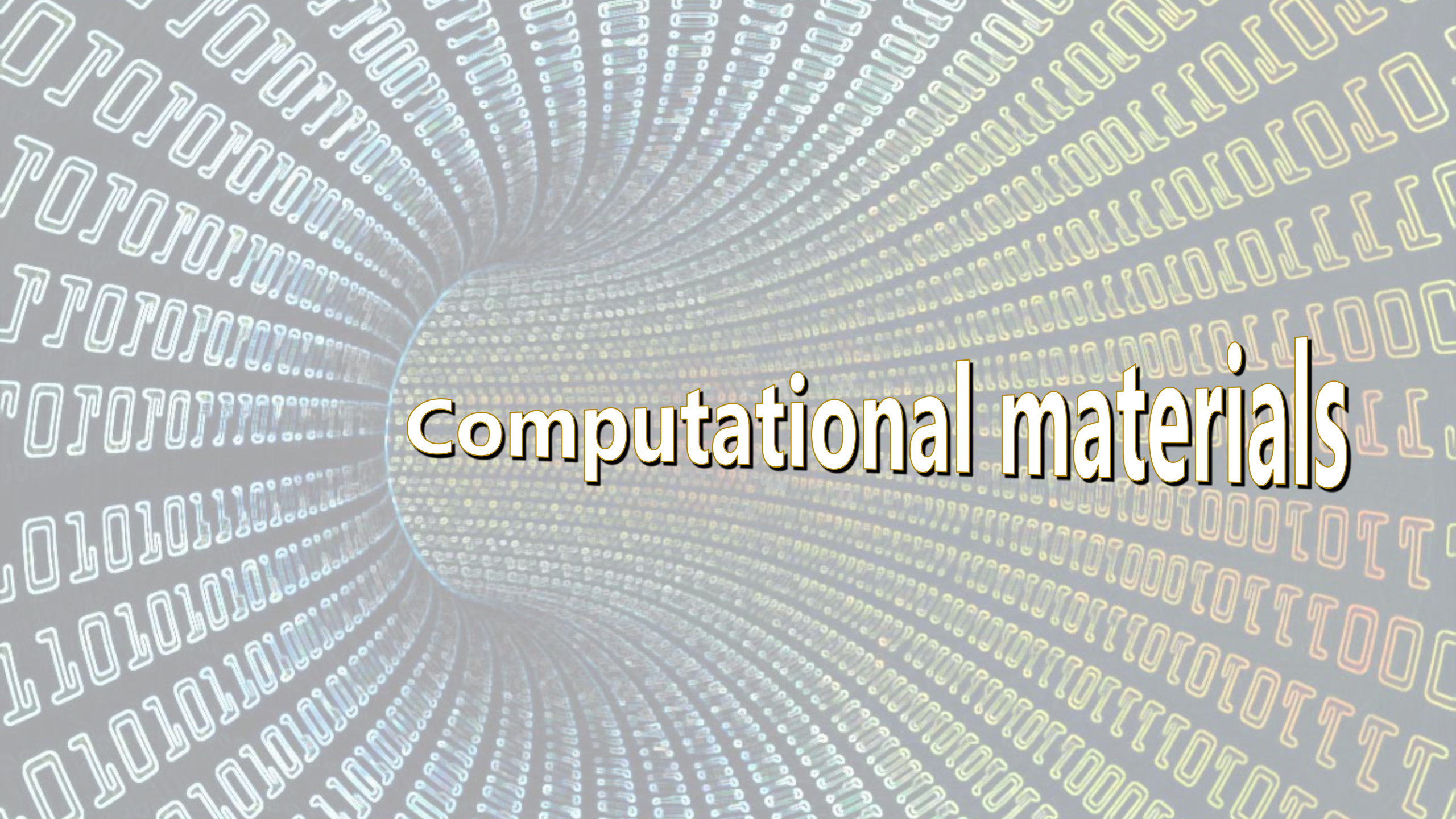


Used to be our favorite material for light bulbs

Too brittle …

www.iter.org

What Frank was talking about …


Computational materials

What do we need to solve?

In principle, the Schrödinger equation …

$$\mathbf{H}\,\psi = \mathbf{E}\,\psi$$

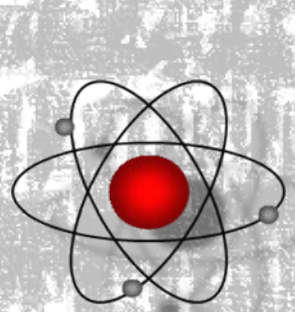
The wavefunction of steel

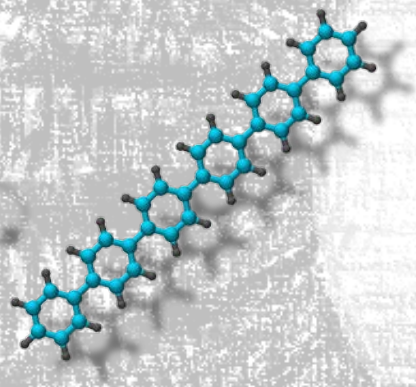
Sandviken, Sweden, 1901, courtesy of Pasha Korzhavyi

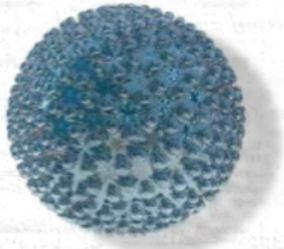
What do we need to solve?

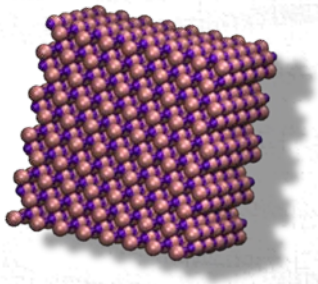
In practice, density-functional theory (DFT) and beyond …

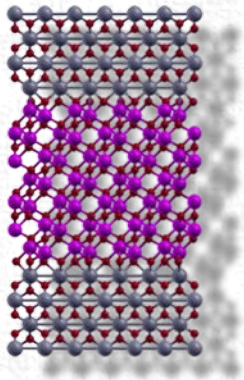
Ab-initio theory for ….

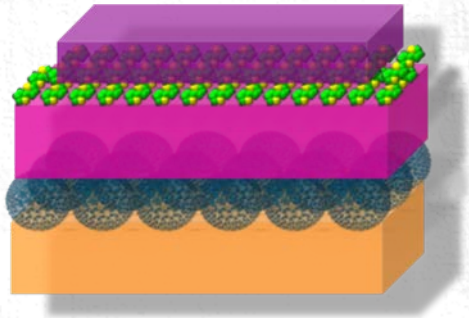












Atoms

Molecules

Clusters

Bulk crystals

Surfaces & interfaces

Nanostructures

Why is DFT tractable?

We replace the intercting many-electron system by an auxiliary system of non-interacting electrons of the same (exact) density.

Kohn-Sham equation

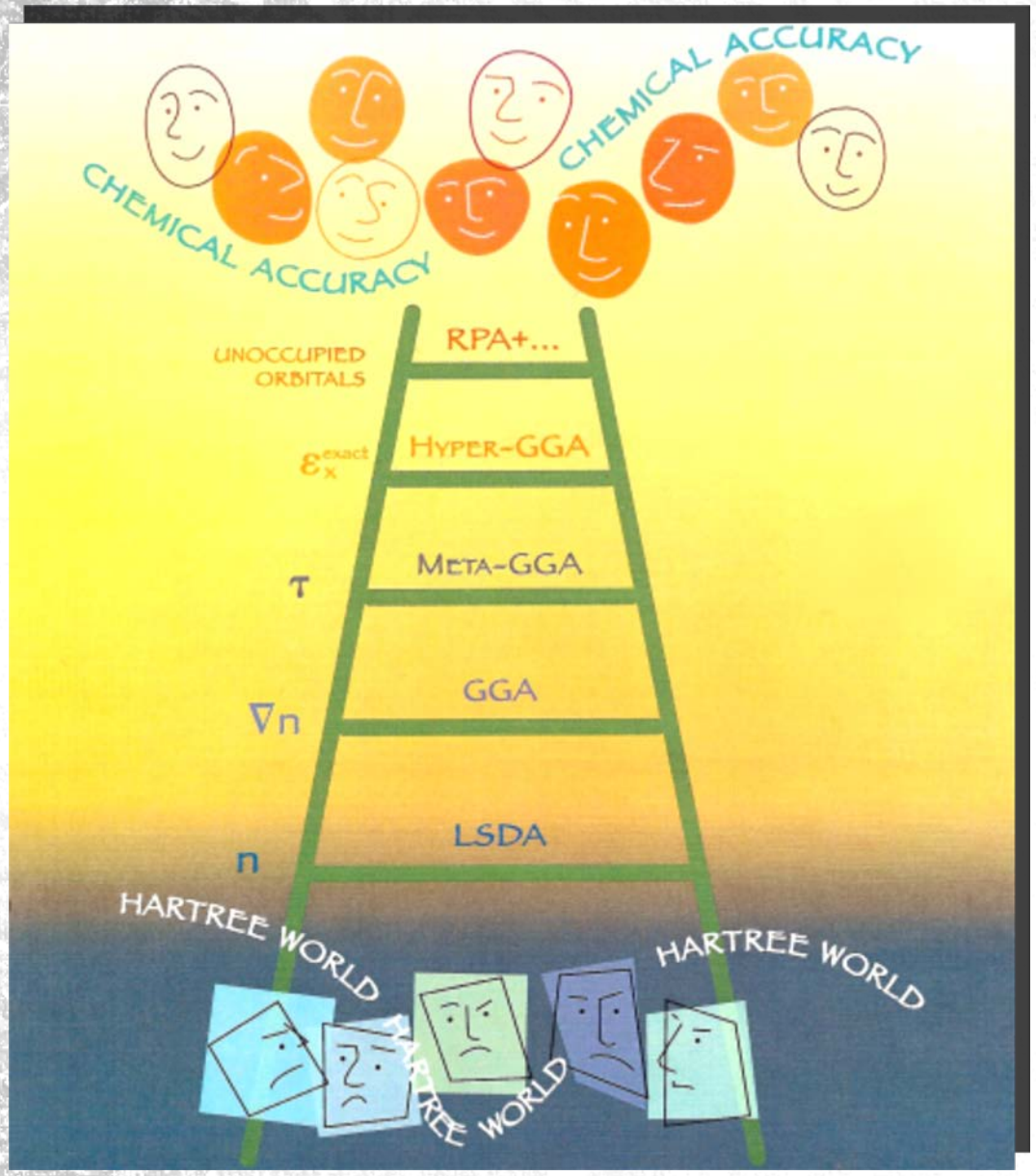
$$\left\{-\frac{1}{2}\nabla^2+V(\mathbf{r})+\int d\mathbf{r}'\sum_j u_j^*(\mathbf{r}')u_j(\mathbf{r}')\frac{1}{|\mathbf{r}-\mathbf{r}'|}+V_{xc}(\mathbf{r})-\epsilon_i\right\}u_i(\mathbf{r})=0$$

$$n(\mathbf{r})=\sum_i^N |u_i(\mathbf{r})|^2$$

You may call it a surrogate model

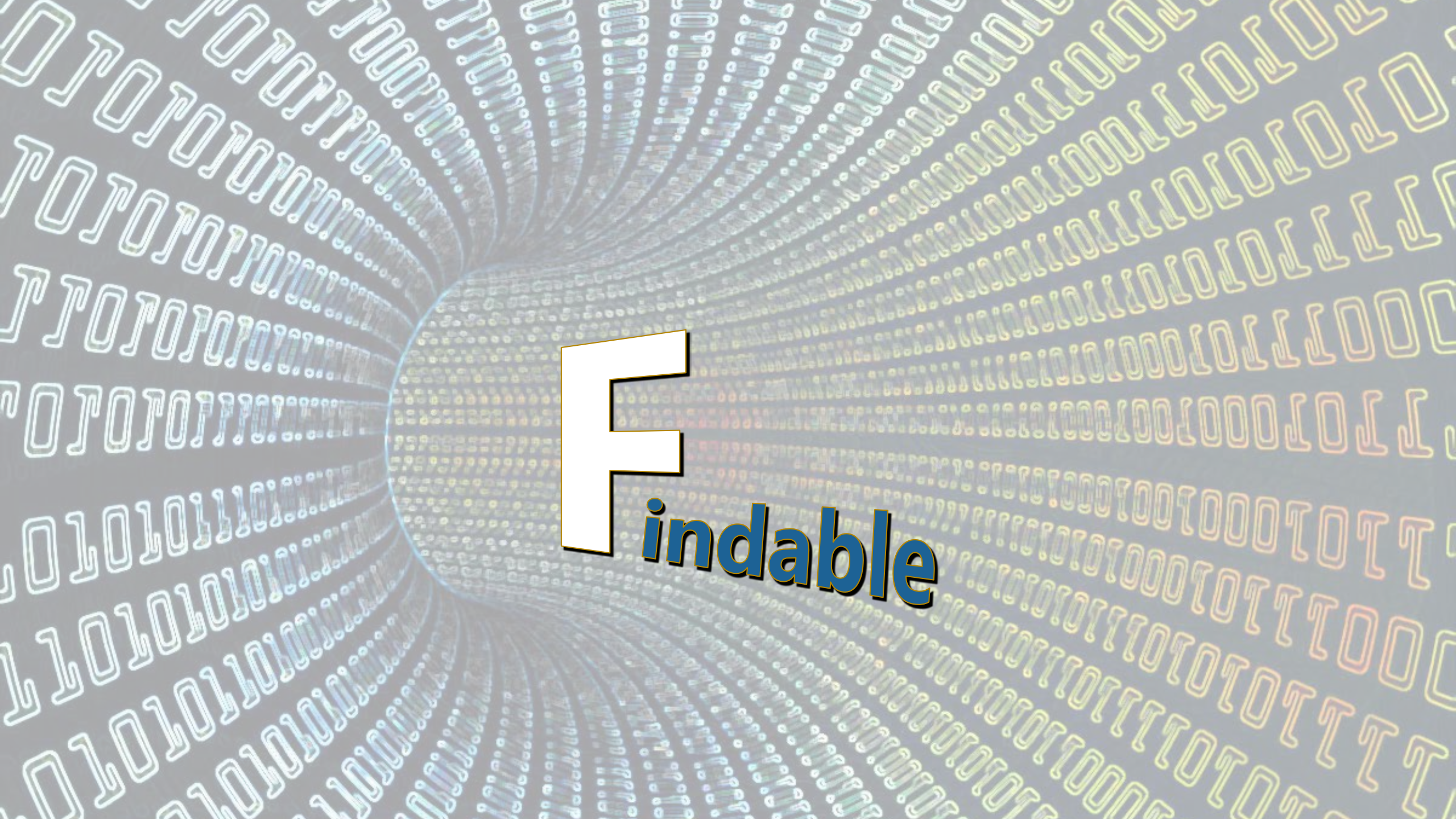
Jacob's ladder of DFT

John Perdew's dream ...

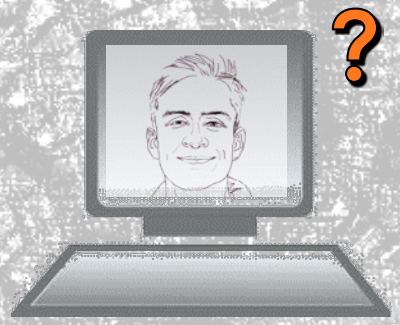


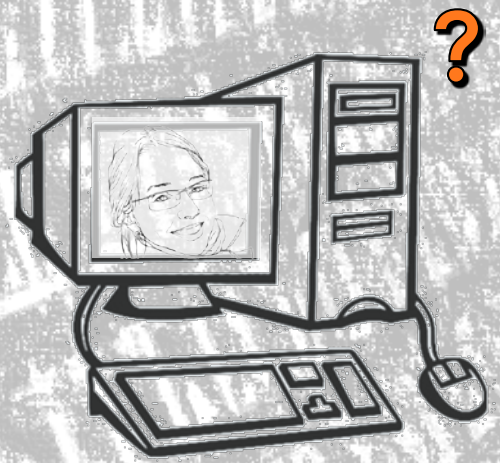
graphics from Kieron Burke

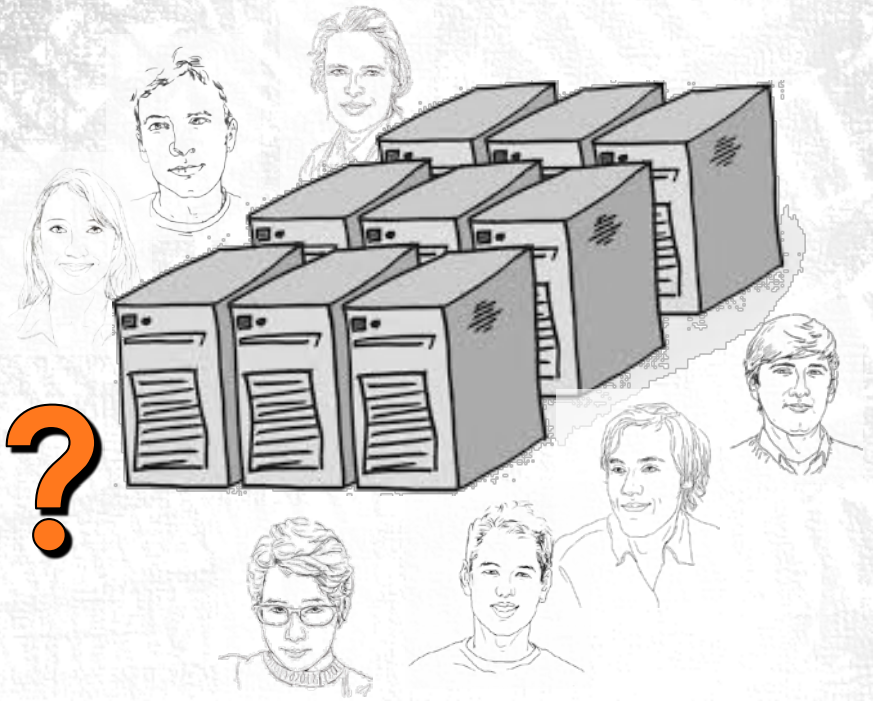

What is FAIR in materials data?

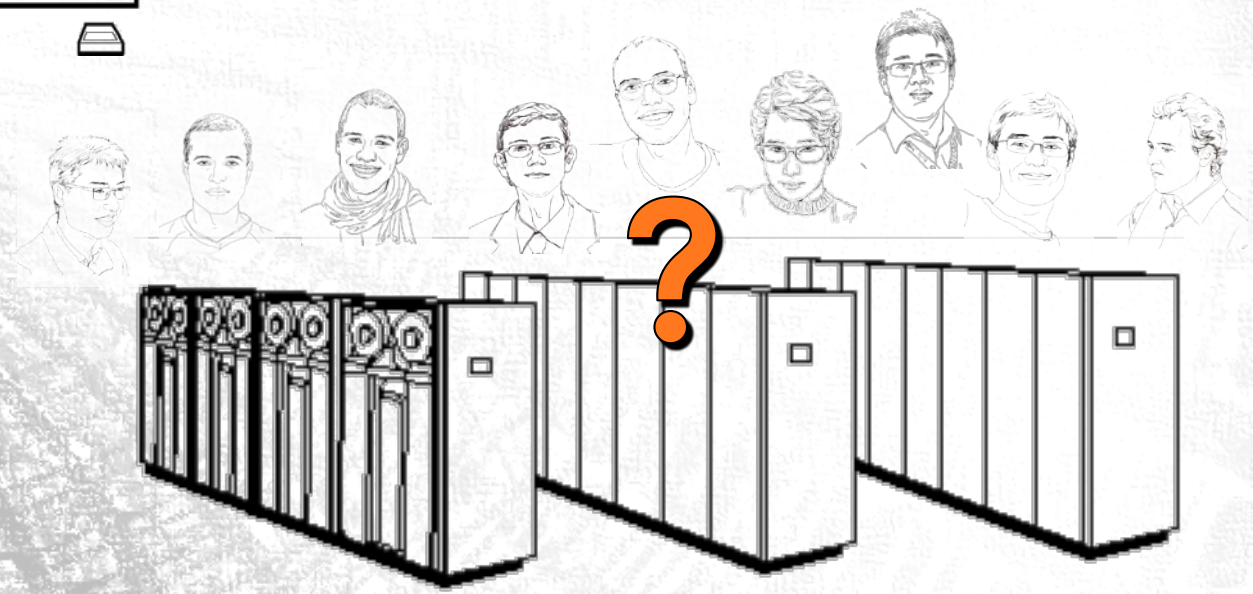

Findable

Findable?

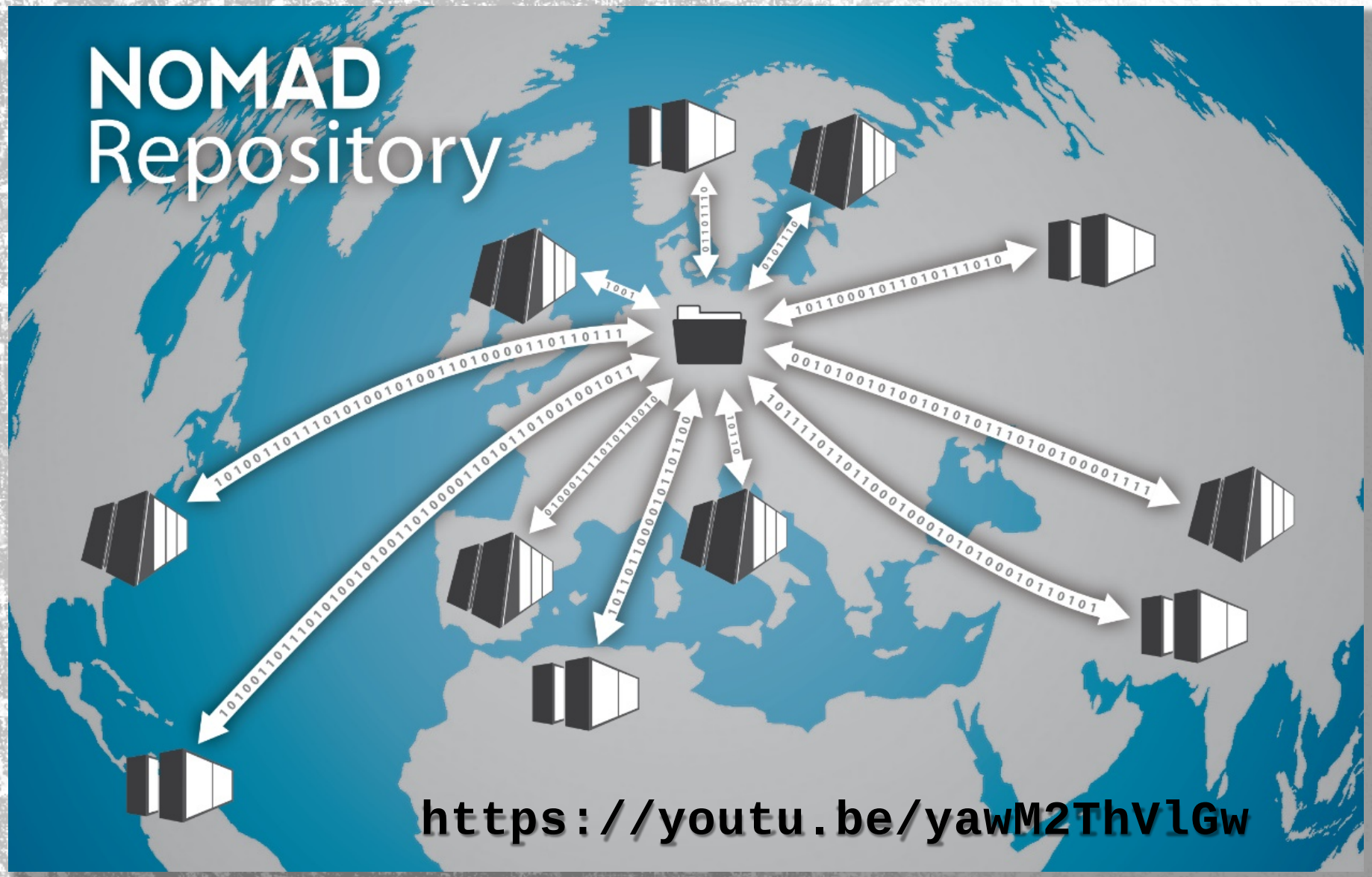


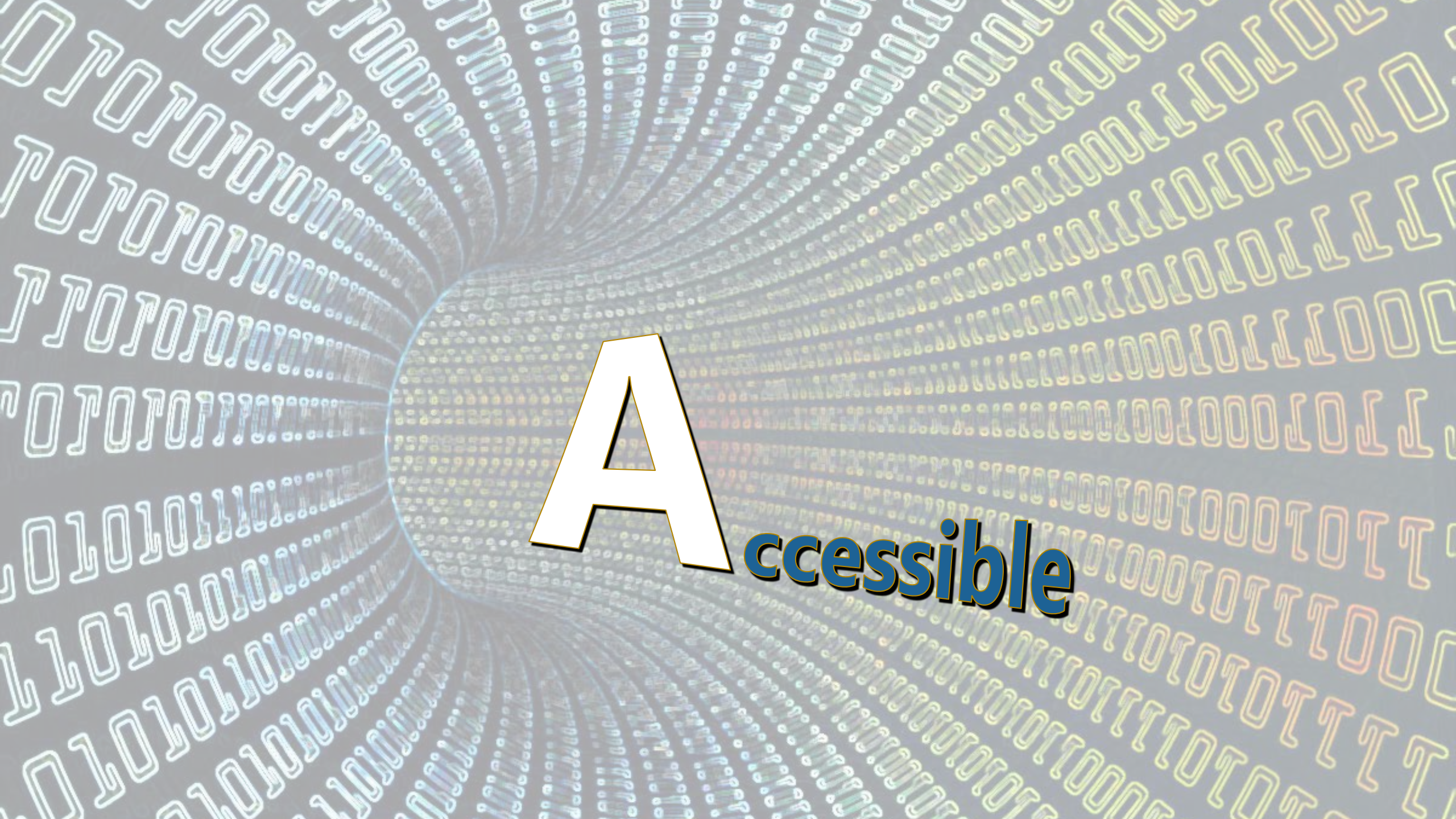




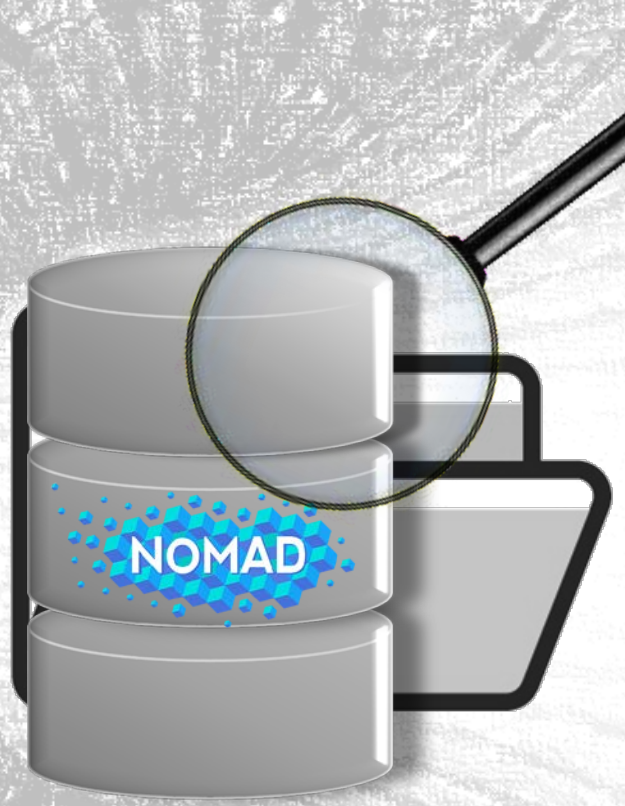


Organizing scientific data ...




Accessible

Accessible?



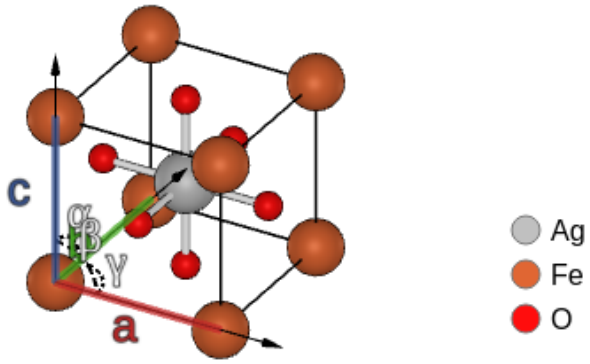
All NOMAD services can also be reached through APIs

What are our services?

The NOMAD Encyclopedia

$AgFeO_3$ - space group 221

Structure



Show axis Show bonds Virtual Reality files

System type: bulk

Space group: 221

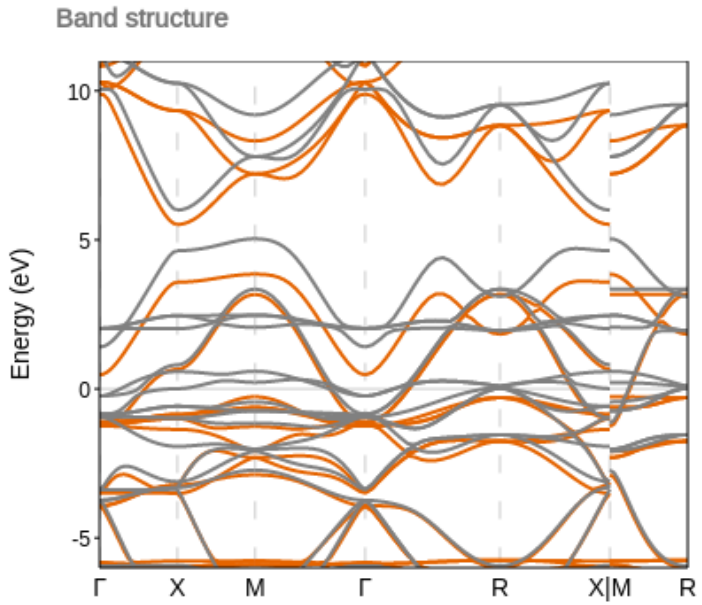
Structure type: CaO3Ti (Cubic Perovskite)

Methodology

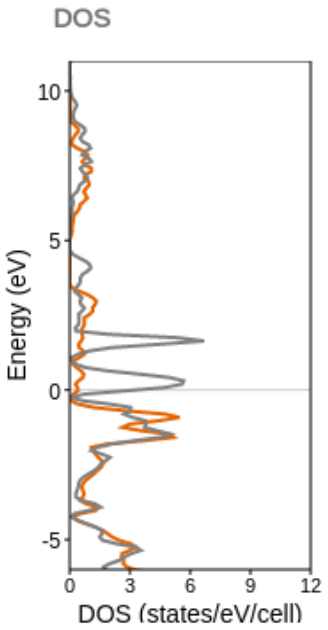
Available calculations

Functional	Code
7 GGA	7 VASP

Electronic structure



From calculation **1648548**
(GGA - VASP)



From calculation **383297**
(GGA - VASP)

Spin ⇧ Spin ⇩

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

Advanced Graphics

Remote visualization

No need for installing software

Virtual-reality tools

From Google cardboard
to full VR environment

https://nomad-coe.eu/the-project/outreach/outreach-videos

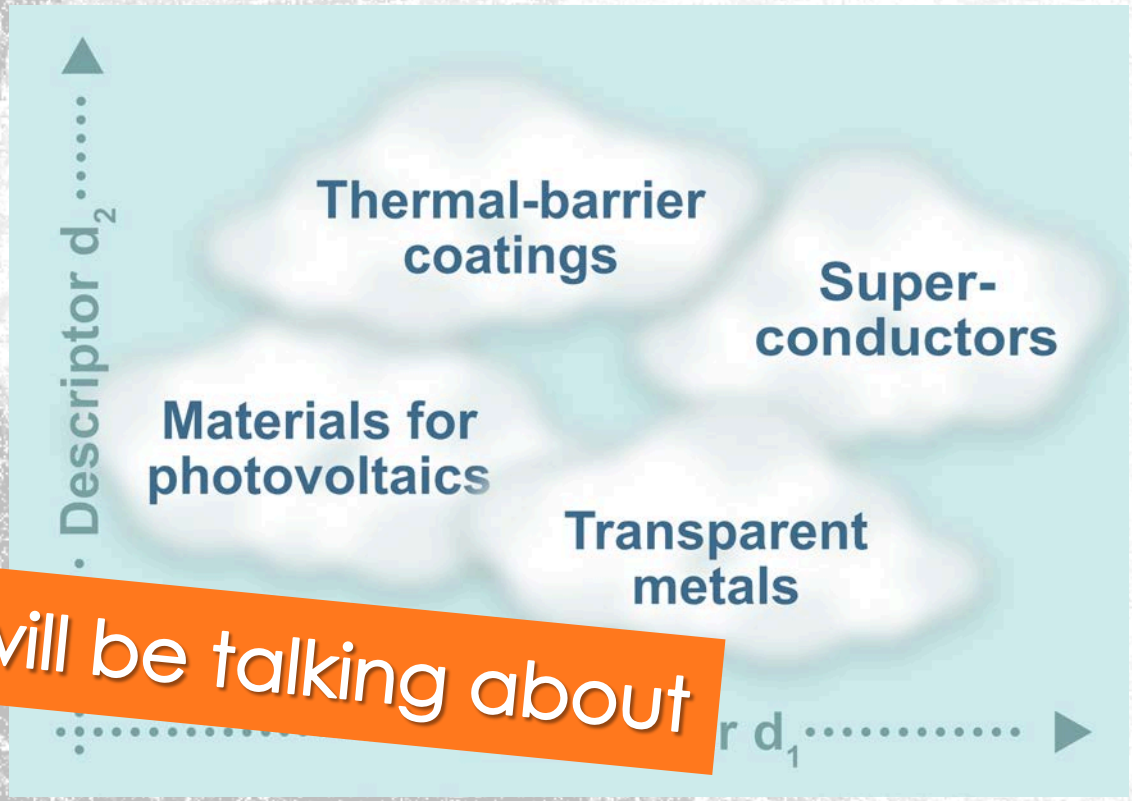
The NOMAD Analytics Toolkit

- Set of ready-to use tutorials on
  - Crystal structure prediction, property prediction, error estimates, classification of materials, and more …
- No need to install any software
- No computational capacity
- No registration

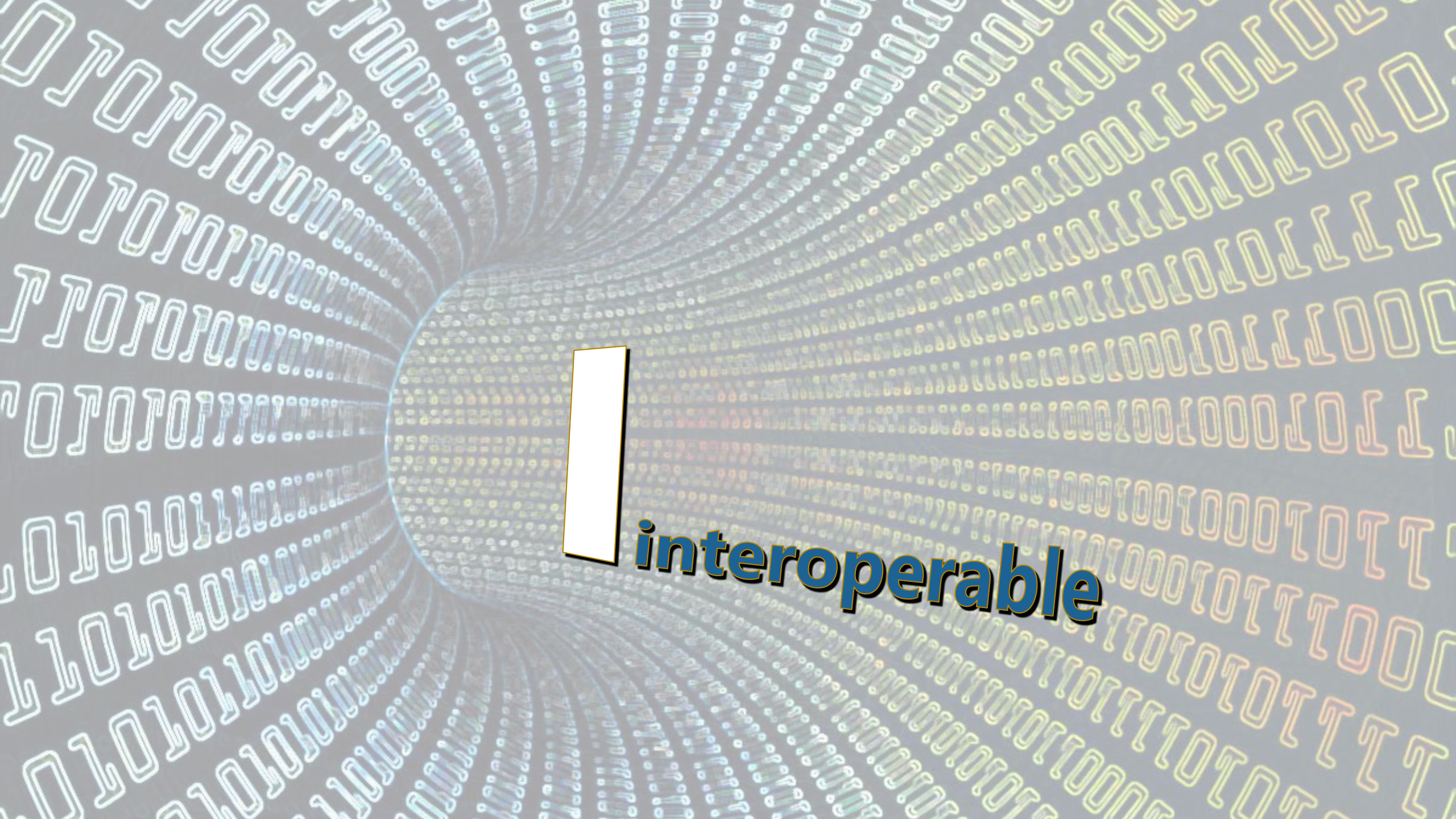
https://analytics-toolkit.nomad-coe.eu/

Our scientific vision is to draw maps

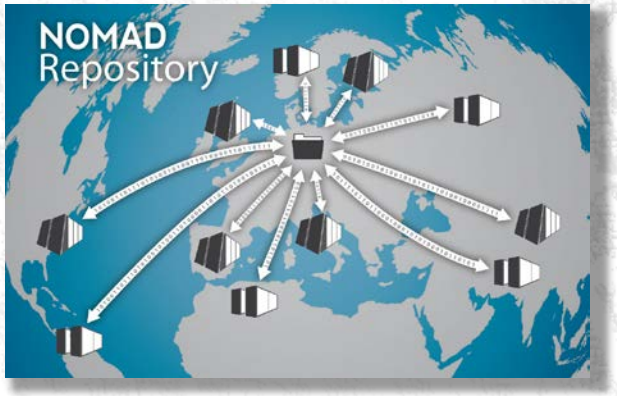
What are the actuators behind the trends and patterns that are invisible to the human eye?



What Matthias will be talking about


interoperable

The NOMAD Archive



More than 50 million calculations coming from …

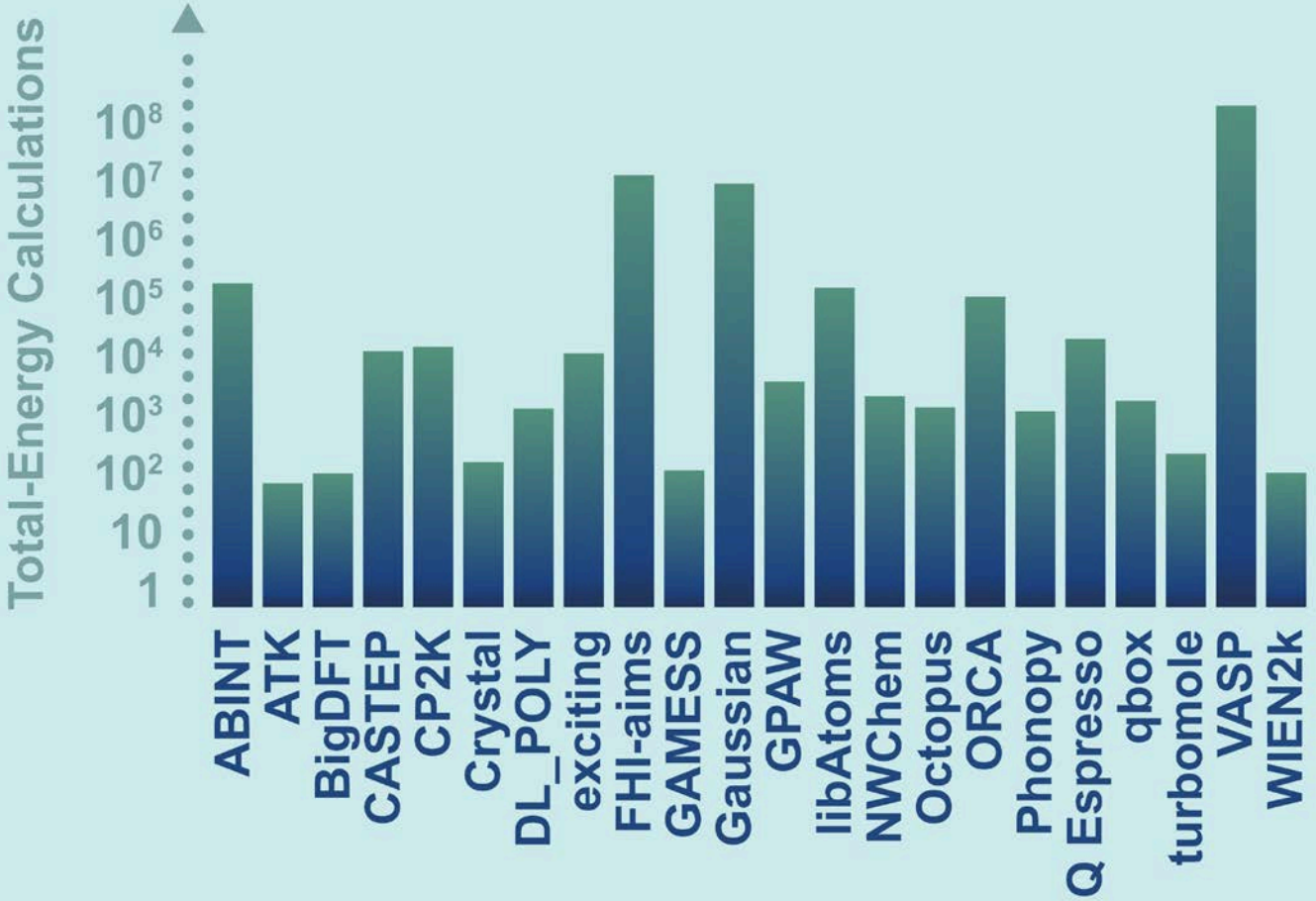
- 40 different codes

Normalized data

- Unified format, units, …

Metadata

- Crucially important
- Unique description of data
  - Every output is fully parsed




Data quality ...

What do these codes do (not) have in common?

Iterative solution of Kohn-Sham equations

$$\left\{-\frac{1}{2}\nabla^2+V(\mathbf{r})+\int d\mathbf{r}'\sum_j u_j^*(\mathbf{r}')u_j(\mathbf{r}')\frac{1}{|\mathbf{r}-\mathbf{r}'|}+V_{xc}(\mathbf{r})-\epsilon_i\right\}u_i(\mathbf{r})=0$$

Expand KS wavefunctions into basis set

$$u_i(\mathbf{r}) = \sum_n c_n\phi_n(\mathbf{r})$$

Solve (generalized) eigenvalue problem

$$\sum_{n'}\left(H_{n,n'}-\varepsilon_n S_{n,n'}\right)c_{n'} = 0$$

Quality depends on number of basis functions

Is it possible to compare results of different codes ?

Science
AAAS

RESEARCH ARTICLE

DFT METHODS

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

Data quality

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] Matt I. J. Probert,[25] Keith Refson,[38,39] Manuel Richter,[31,32] Gian-Marco Rignanese,[19] Santanu Saha,[20] Matthias Scheffler,[15,40] Martin Schlipf,[22] Karlheinz Schwarz,[5] Sangeeta Sharma,[12] Francesca Tavazza,[17] Patrik Thunström,[41] Alexandre Tkatchenko,[15,42] Marc Torrent,[28] David Vanderbilt,[23] Michiel J. van Setten,[19] Veronique Van Speybroeck,[1] John M. Wills,[43] Jonathan R. Yates,[29] Guo-Xu Zhang,[44] Stefaan Cottenier[1,45]*

Delta test

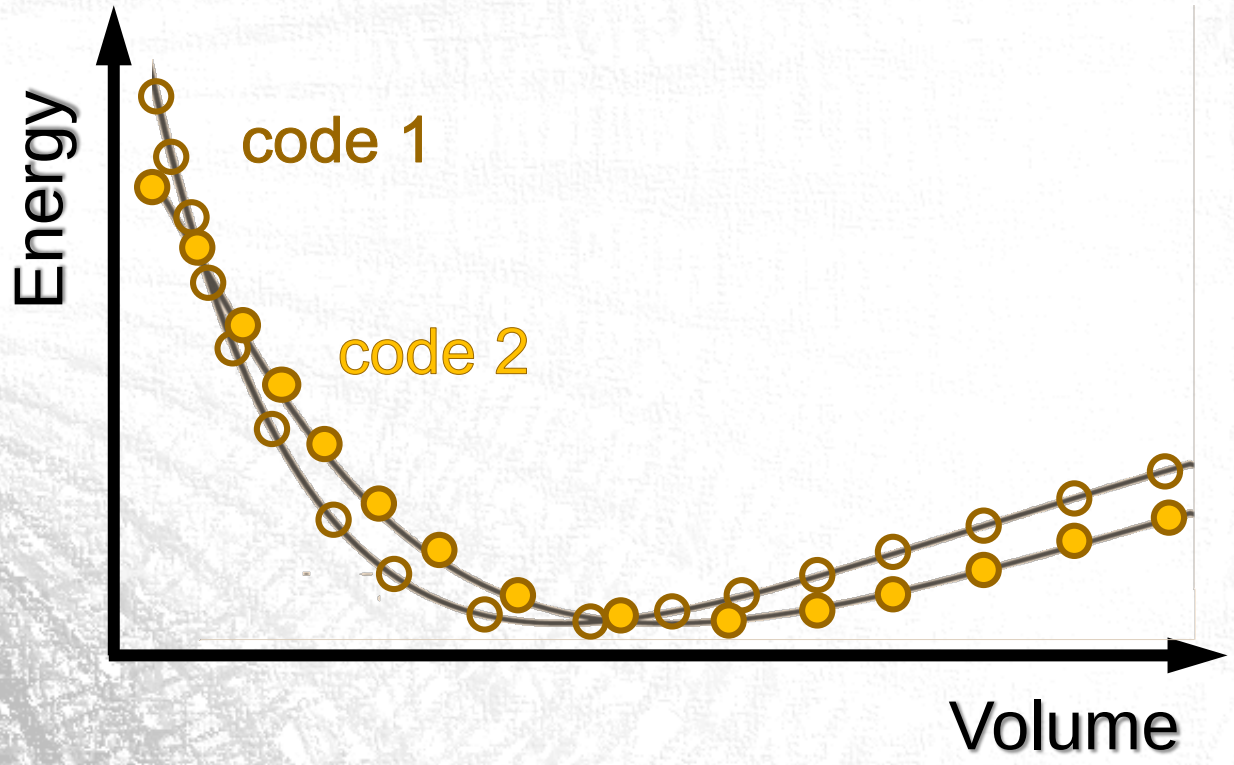
Compute E(V) using PBE

Do the same with other code

Fit to Burch-Murnaghan equation of state

Quality factor

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



Delta test

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 test

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 test

Agreement for save numerical setting

Excellent overall agreement between codes – great!

... but just the beginning

What about ...

... other systems?

Surfaces, defects, molecules, ...

... other quantities?

Band gaps, barriers, spectra, ...

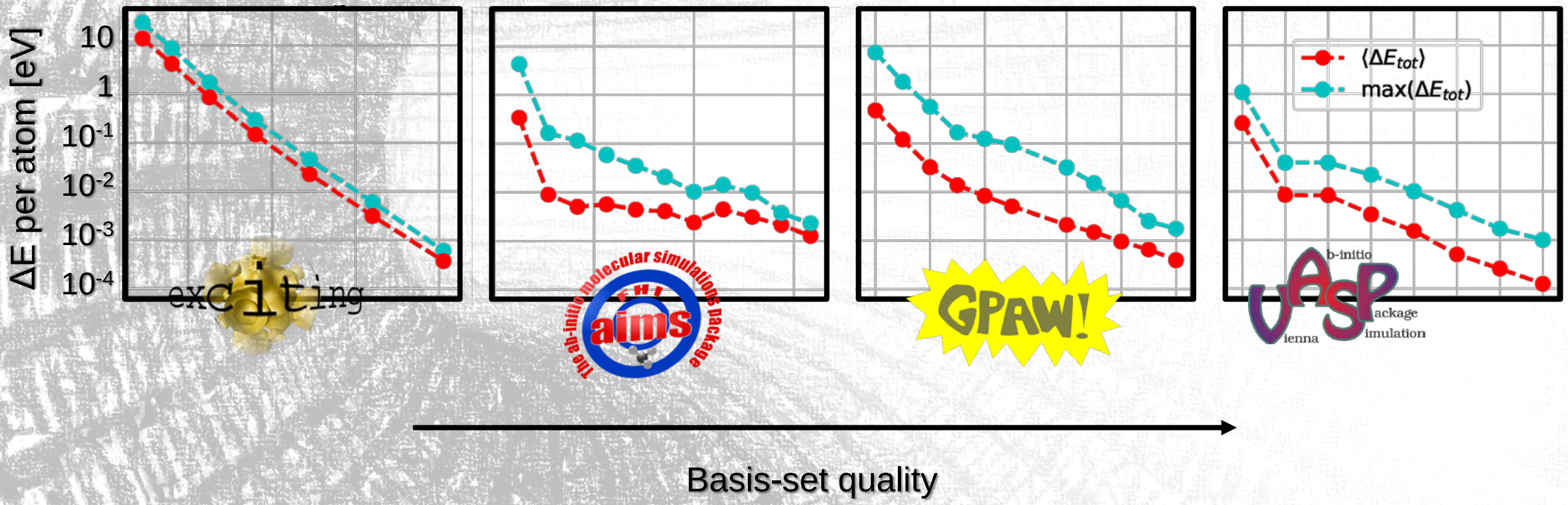
Larger discrepancies for certain elements

O_2

	exciting	FHI-AIMS/tier2	FLEUR	FPLO/T+F+s	RSPt	WIEN2k/acc	VASP2011/VASP	VASP2012/VASP
exciting		0.1	0.4	1.8	0.7	0.6	7.7	0.2
FHI-AIMS/tier2	0.1		0.3	1.7	0.6	0.5	7.6	0.1
FLEUR	0.4	0.3		1.4	0.3	0.3	7.2	0.2
FPLO/T+F+s	1.8	1.7	1.4		1.1	1.2	5.8	1.6
RSPt	0.7	0.6	0.3	1.1		0.1	6.9	0.5
WIEN2k/acc	0.6	0.5	0.3	1.2	0.1		7.1	0.4
VASP2011/VASP	7.7	7.6	7.2	5.8	6.9	7.1		7.5
VASP2012/VASP	0.2	0.1	0.2	1.6	0.5	0.4	7.5	

Numerical quality control in DFT calculations

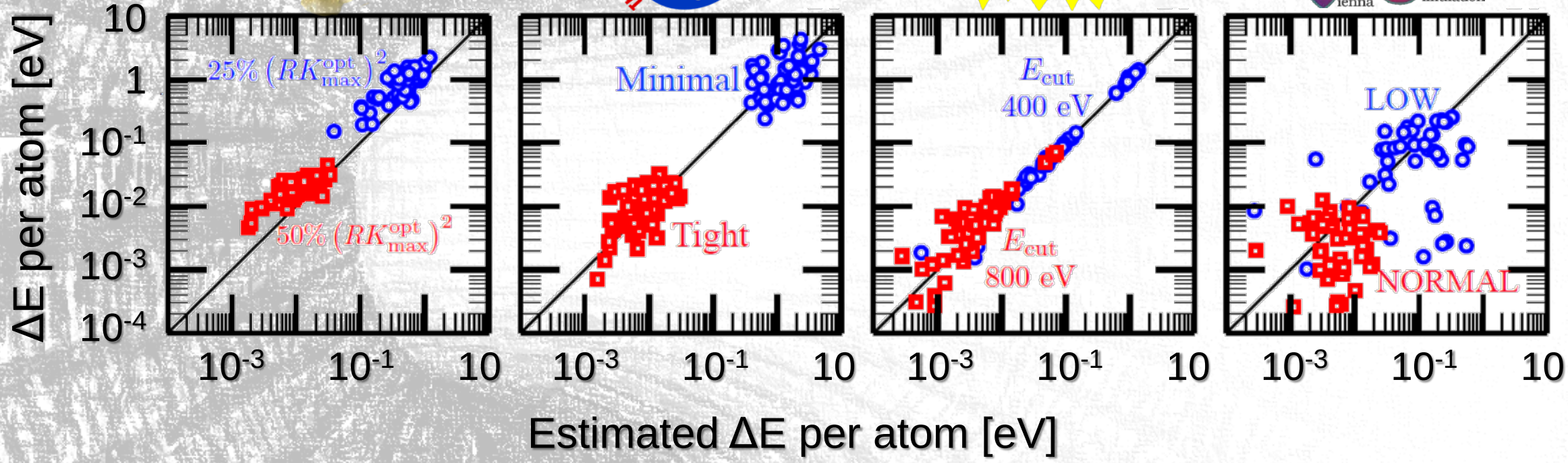
Total energies of 71 elemental solids



Error prediction based on errors for elements

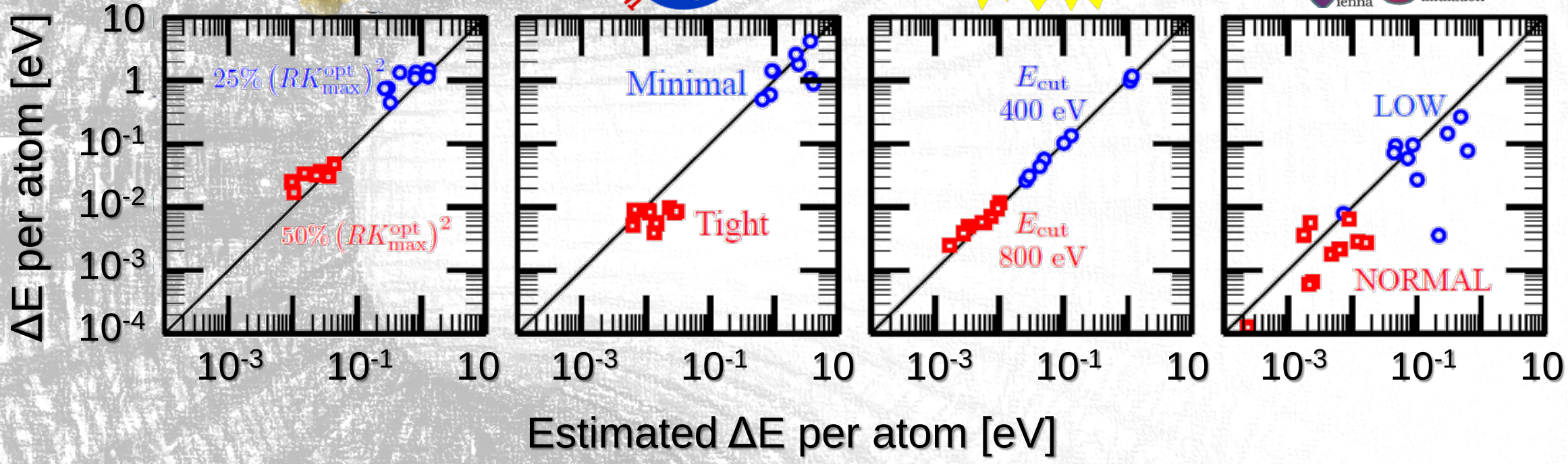
$$\widehat{\Delta E} = \frac{1}{N}\sum_{I} N_I \Delta E_I$$

Error prediction for 63 binaries

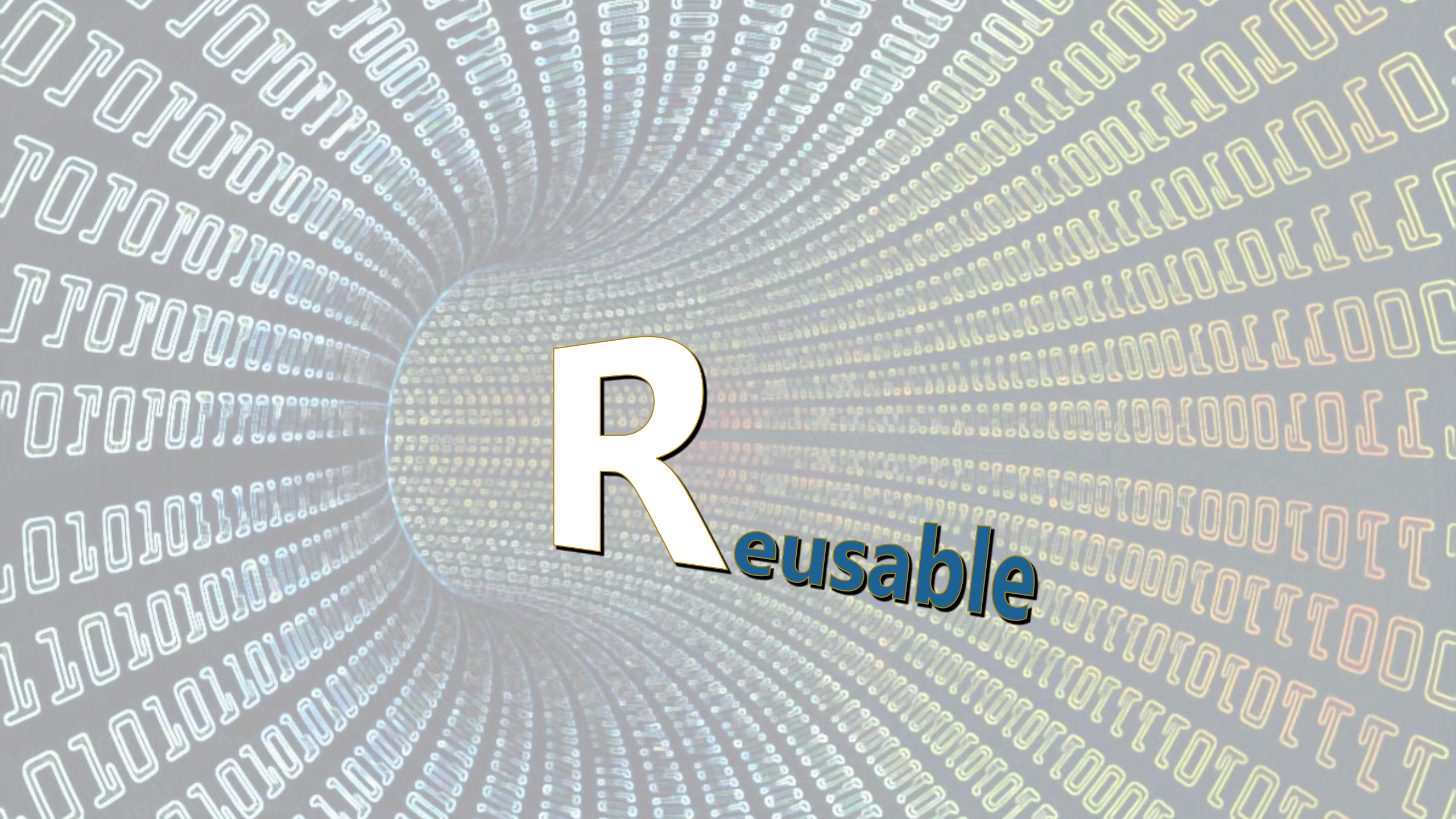


C. Carbogno, et al., preprint.

Error prediction for 10 arbitrary ternaries



Predictive power enables full interoperability


Reusable

One material – many possible applications

Reusable = repurposeable

Tungsten and its alloys

- Light bulbs
- Fusion reactors
- Radiation shielding

Example: TiO_2

- Support material for heterogeneous catalysis
- Photovoltaics - in dye-sensitized solar cells
- White color in paints and sunscreens

Solving the quantum-mechanical problem, the results (data) contain much more information than published !!

NOvel MAterials Discovery

The FAIR Concept for Big-Data-Driven Materials Science

C. Draxl and M. Scheffler, MRS Bulletin 43, 676(2018).

Findable
Accessible
Interoperable
Re-usable

51,786,061 open-access calculations

Mark D. Wilkinson, ... P. Wittenburg, ... et al.
The FAIR Guiding Principles for scientific data management and stewardship
Sci. Data **3**, 160018 (2016).


Yet another database?

Other materials databases – and their purpose

Materials Project

https://materialsproject.org

Batteries, phase diagrams, etc.

Mainly 1 code: VASP

Particular functional: PBE (+U)

Not open for uploads

AFLOW

http://aflowlib.org

1,859,011 compounds, > 184,042,089 calculated properties

1 code: VASP

The Open Quantum Materials Database (OQMD)

http://oqmd.org

DFT-calculated thermodynamic and structural properties

1 code: VASP

The NOMAD Repository is orthorgonal

All input and output files of more than 50 million calculations
- Raw data
- Data quality *known*

Handling requires only few metadata
- Authors, code & version, upload date, …

Hosted by Max-Planck Computing and Data Facility
- Clones being built

Recommended by Scientific Data

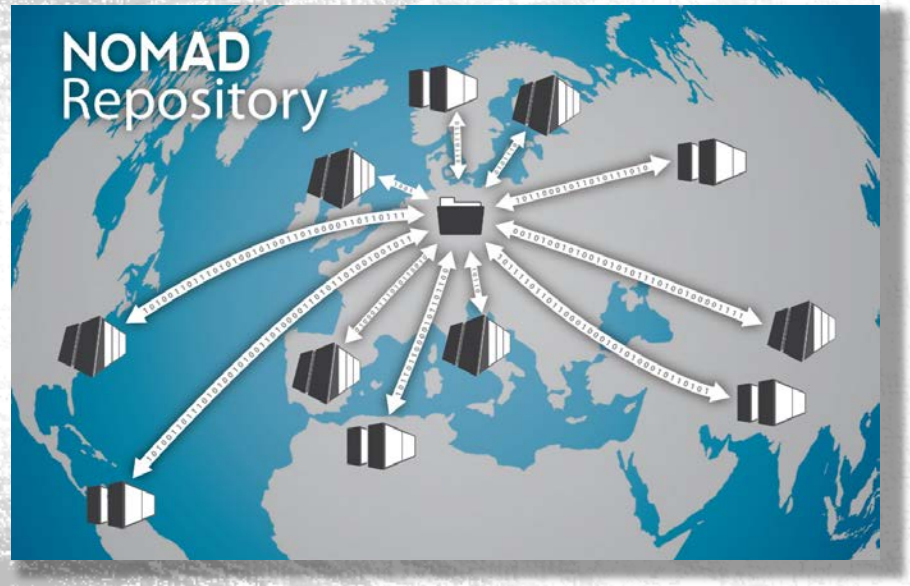
Worldwide largest collection
- Contains data of US DBs – raw data only available

NOMAD makes these DBs FAIR(er)


Next steps … … …

From the beginning …

Matthias Scheffler
FHI Berlin



Claudia Draxl
HU Berlin

Ciaran Clissman
Pintail Dublin

Angel Rubio
MPSD Hamburg

Kristian Thygesen
DTU Lyngby

Kimmo Koski
CSC Helsinki

Risto Nieminen
Aalto Univ. Helsinki

Claudia Draxl
HU Berlin

Matthias Scheffler
FHI Berlin

Daan Frenkel
Univ. Cambridge

†

Alessandro De Vita
Kings College London

†

Stefan Heinzel
MPSCD Garching

Dieter Kranzlmüller
LRZ Munich

Francesc Illas
Univ. Barcelona

Jose Maria Cela
BSC Barcelona

Funding from the European Union's Horizon 2020 research and innovation programme, grant agreement No 676580.

FAIRDI

FAIR Data Infrastructure for Physics, Chemistry, Materials Science, and Astronomy

Non-profit association just being founded

Five pillars

Pillar A - Computational materials science (NOMAD)

Pillar B - Experimental materials science

Pillar C - Biophysical and soft-matter simulations

Pillar D - Astronomy and space-situational awareness

Pillar E - User management, intellectual property rights, cyber security

horizontal

https://fairdi.eu


Thanks !!