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
- Biocompatible
- Light-weight & ductile
- Heat-resistant
- Hard & tough
- High quantum yield
- Energy-saving
- Desired color
Example: fusion reactors

Tungsten?

Highest melting point: 3683K

Used to be our favorite material for light bulbs

Too brittle ...

What Frank was talking about ...
Computational materials
What do we need to solve?

In principle, the Schrödinger equation ...

\[ H \psi = E \psi \]
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

Clusters

Surfaces & interfaces

Molecules

Bulk crystals

Nanostructures
Why is DFT tractable?

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

Kohn-Sham equation

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

\[
n(r) = \sum_{i=1}^{N} |u_i(r)|^2
\]

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?
Organizing scientific data...

NOMAD Repository

https://youtu.be/yawM2ThVlGw
Accessible
Accessible?

All NOMAD services can also be reached through APIs. What are our services?
The NOMAD Encyclopedia

AgFeO₃ - space group 221

Structure

Electronic structure

- Band structure
  - Energy (eV)
  - From calculation 1648548 (GGA - VASP)
  - Spin ↑ Spin ↓
- DOS
  - Energy (eV)
  - From calculation 383297 (GGA - VASP)

System type: bulk
Space group: 221
Structure type: CaO₃Ti (Cubic Perovskite)

Methodology

Available calculations
- Functional: 7 GGA
- Code: 7 VASP

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

Iterative solution of Kohn-Sham equations

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

Expand KS wavefunctions into basis set

$$u_i(r) = \sum_n c_n \phi_n(r)$$

Solve (generalized) eigenvalue problem

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

Quality depends on number of basis functions

Is it possible to compare results of different codes?
Reproducibility in density functional theory calculations of solids

**Delta test**

Compute $E(V)$ using PBE
Do the same with other code
Fit to Burch-Murnaghan equation of state

**Quality factor**

$$\Delta = \left\langle \sqrt{\frac{\int \Delta E^2(V) dV}{\Delta V}} \right\rangle$$
<table>
<thead>
<tr>
<th>Code</th>
<th>Version</th>
<th>Basis</th>
<th>Electron treatment</th>
<th>Δ-value</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exciting</td>
<td>development LAPW+xlo version</td>
<td>all-electron</td>
<td>0 meV/atom</td>
<td>Exciting [10,16]</td>
<td></td>
</tr>
<tr>
<td>FHI-aims</td>
<td>081213</td>
<td>tier2 numerical all-electron (relativistic atomic_zora scalar)</td>
<td>0.1 meV/atom</td>
<td>ASE [2,16]</td>
<td></td>
</tr>
<tr>
<td>WIEN2k</td>
<td>13.1</td>
<td>LAPW/APW+lo all-electron</td>
<td>0.2 meV/atom</td>
<td>S. Cottenier [16]</td>
<td></td>
</tr>
<tr>
<td>FHI-aims</td>
<td>081213</td>
<td>tier2 numerical all-electron (relativistic zora scalar 1e-12)</td>
<td>0.3 meV/atom</td>
<td>ASE [2]</td>
<td></td>
</tr>
<tr>
<td>Quantum ESPRESSO</td>
<td>5.1</td>
<td>plane waves</td>
<td>SSSP Accuracy (mixed NC/US/PAW potential library)</td>
<td>0.3 meV/atom [12,16]</td>
<td></td>
</tr>
</tbody>
</table>

https://molmod.ugent.be/deltacodesdft
<table>
<thead>
<tr>
<th>Code</th>
<th>Version</th>
<th>Basis</th>
<th>Electron treatment</th>
<th>Δ-value</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exciting</strong></td>
<td>development LAPW+xlo version</td>
<td>all-electron</td>
<td>0 meV/atom</td>
<td></td>
<td>Exciting [10,16]</td>
</tr>
<tr>
<td><strong>FHI-aims</strong></td>
<td>081213</td>
<td>tier2 numerical</td>
<td>all-electron (relativistic atomic_zora scalar)</td>
<td>0.1 meV/atom</td>
<td>ASE [2,16]</td>
</tr>
<tr>
<td><strong>WIEN2k</strong></td>
<td>13.1</td>
<td>LAPW/APW+lo</td>
<td>all-electron</td>
<td>0.2 meV/atom</td>
<td>S. Cottenier [16]</td>
</tr>
<tr>
<td><strong>FHI-aims</strong></td>
<td>081213</td>
<td>tier2 numerical</td>
<td>all-electron (relativistic zora scalar 1e-12)</td>
<td>0.3 meV/atom</td>
<td>ASE [2]</td>
</tr>
<tr>
<td>Quantum ESPRESSO</td>
<td>5.1</td>
<td>plane waves</td>
<td>SSSP Accuracy (mixed NC/US/PAW potential library)</td>
<td>0.3 meV/atom</td>
<td>QuantumESPRESSO [12,16]</td>
</tr>
</tbody>
</table>
Delta test

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

<table>
<thead>
<tr>
<th></th>
<th>exciting</th>
<th>FHI-AIMS/tier2</th>
<th>FLEUR</th>
<th>FPLO/T+F+s</th>
<th>RSpt</th>
<th>WIEN2k/acc</th>
<th>VASP2011/VASP</th>
<th>VASP2012/VASP</th>
</tr>
</thead>
<tbody>
<tr>
<td>exciting</td>
<td>0.1</td>
<td>0.1</td>
<td>0.3</td>
<td>1.8</td>
<td>0.7</td>
<td>0.6</td>
<td>7.7</td>
<td>0.2</td>
</tr>
<tr>
<td>FHI-AIMS/tier2</td>
<td>0.4</td>
<td>0.4</td>
<td>0.3</td>
<td>1.7</td>
<td>0.6</td>
<td>0.5</td>
<td>7.6</td>
<td>0.1</td>
</tr>
<tr>
<td>FLEUR</td>
<td>0.4</td>
<td>0.4</td>
<td>0.3</td>
<td>1.4</td>
<td>0.3</td>
<td>0.3</td>
<td>7.2</td>
<td>0.2</td>
</tr>
<tr>
<td>FPLO/T+F+s</td>
<td>1.8</td>
<td>1.7</td>
<td>1.4</td>
<td>1.1</td>
<td>1.2</td>
<td>1.2</td>
<td>5.8</td>
<td>1.6</td>
</tr>
<tr>
<td>RSpt</td>
<td>0.7</td>
<td>0.6</td>
<td>0.3</td>
<td>1.1</td>
<td>0.6</td>
<td>0.1</td>
<td>6.9</td>
<td>0.5</td>
</tr>
<tr>
<td>WIEN2k/acc</td>
<td>0.6</td>
<td>0.5</td>
<td>0.3</td>
<td>1.2</td>
<td>0.1</td>
<td>0.1</td>
<td>7.1</td>
<td>0.4</td>
</tr>
<tr>
<td>VASP2011/VASP</td>
<td>7.7</td>
<td>7.6</td>
<td>7.2</td>
<td>5.8</td>
<td>6.9</td>
<td>7.1</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td>VASP2012/VASP</td>
<td>0.2</td>
<td>0.1</td>
<td>0.2</td>
<td>1.6</td>
<td>0.5</td>
<td>0.4</td>
<td>7.5</td>
<td></td>
</tr>
</tbody>
</table>
Numerical quality control in DFT calculations

Total energies of 71 elemental solids

$\Delta E_{\text{per atom}} [\text{eV}]$

Basis-set quality
Error prediction based on errors for elements

\[ \hat{\Delta E} = \frac{1}{N} \sum_{i} N_i \Delta E_i \]
Error prediction for 63 binaries

Estimated $\Delta E$ per atom [eV]

Error prediction for 10 arbitrary ternaries

Estimated $\Delta E$ per atom [eV]

Predictive power enables full interoperability
One material – many possible applications

Tungsten and its alloys
- Light bulbs
- Fusion reactors
- Radiation shielding

Example: TiO₂
- 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!!
The FAIR Concept for Big-Data-Driven Materials Science


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
Yet another database?
Other materials databases – and their purpose

Materials Project

- Batteries, phase diagrams, etc.
- Mainly 1 code: VASP
- Particular functional: PBE (+U)
- Not open for uploads

AFLOW

- 1,859,011 compounds, > 184,042,089 calculated properties
- 1 code: VASP

The Open Quantum Materials Database (OQMD)

- DFT-calculated thermodynamic and structural properties
- 1 code: VASP
The NOMAD Repository is orthogonal

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 through

NOMAD makes these DBs FAIR(er)
Next steps...
From the beginning ...
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

https://fairdi.eu
Thanks!!