

Making Data and Data Workflows FAIR and Pervasive with nanoHUB: Present and Future

1 M visitors
22,000+ simulation users
per year



Alejandro (Ale) Strachan – strachan@purdue.edu
Reilly Professor of Materials Engineering
Co-Director, nanoHUB
Purdue University





nanoHUB: making simulation & data pervasive



Science codes, data
tools & repos



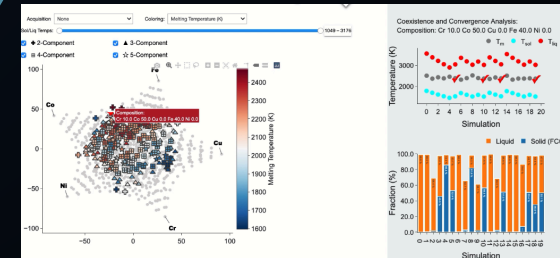
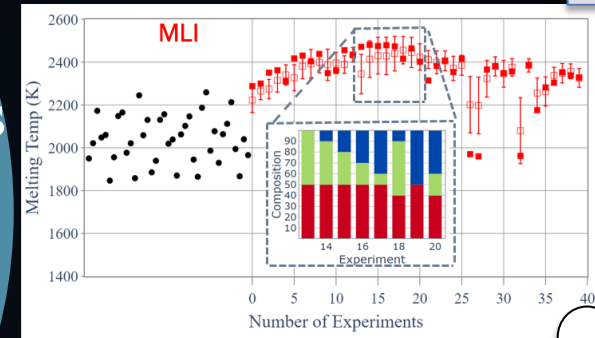
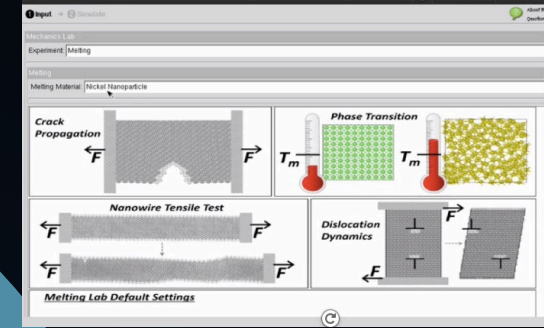
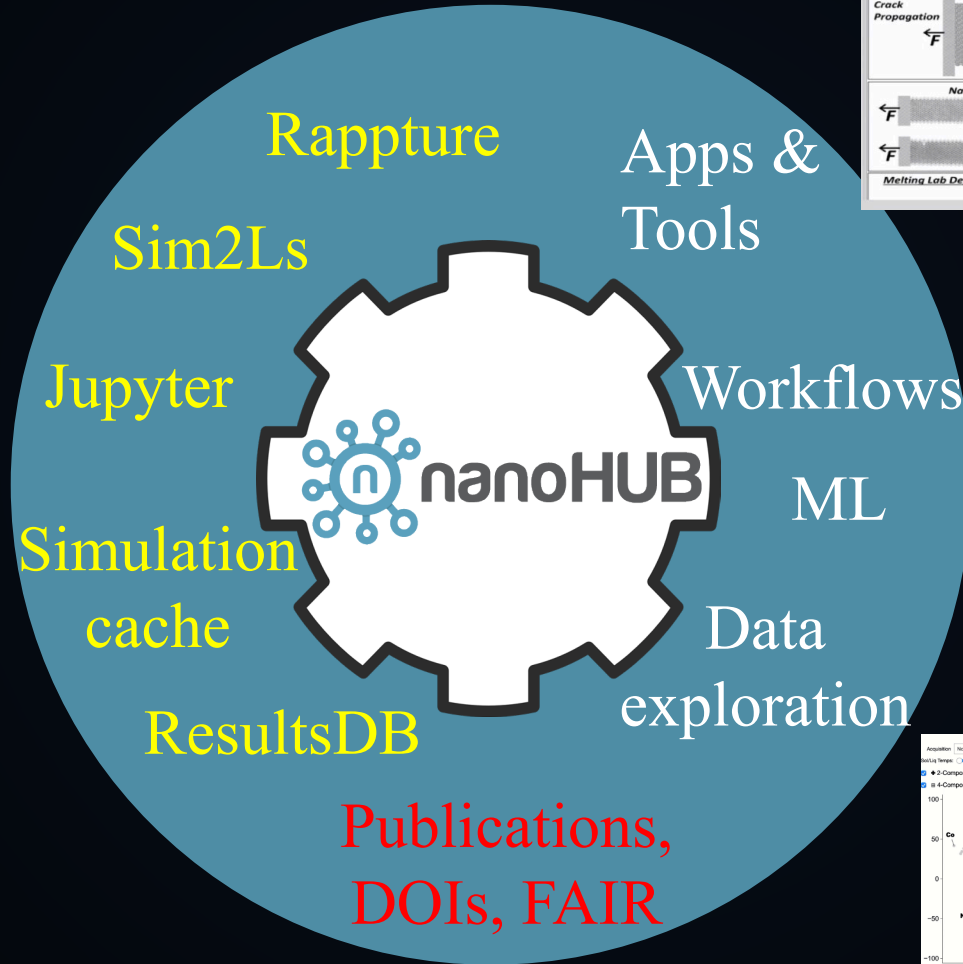
Padre



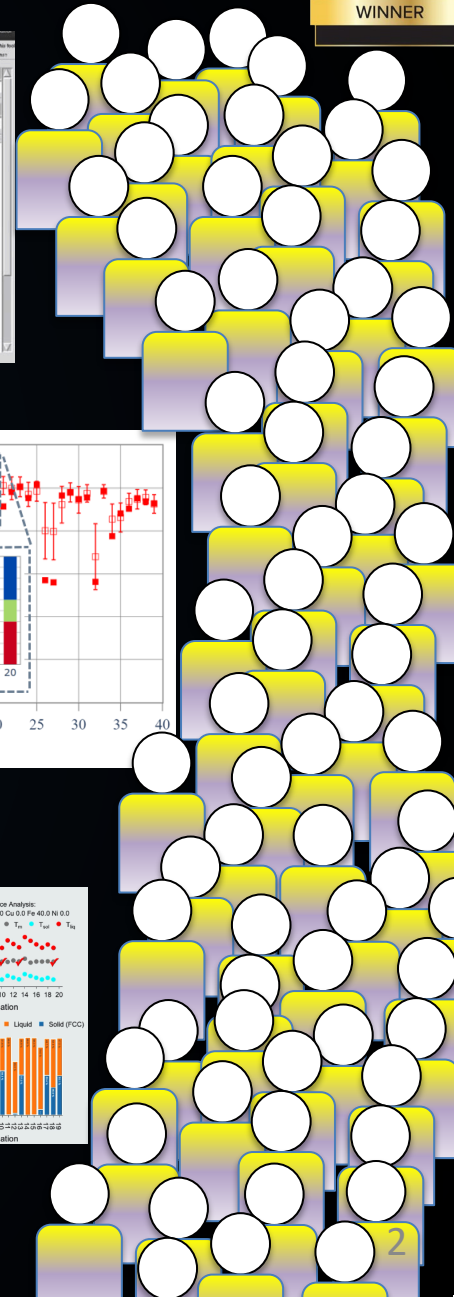
Polymer Modeler



SILVACO



Courses, tutorials, workshops



Access: Turning tools into Apps

```
Projects — vi .//Semicond/MoTe2Crack/Crack/monolayer-MoTe2-hex.in — 104x45

units      real
atom_style full
boundary   p p s

box
region
create_box
lattice

box prism 0 3.5562 0 3.0797 0 9.0000 -1.7781 0.0 0.0
2 box
custom 1.0 a1 3.5562 0.0 0.0 a2 1.7781 3.0797 0.0 a3 0.0 0.0 9.0 &
basis 0.0 0.0 0.0

create_atoms 1 single 0.0 0.0 0.5 units lattice
create_atoms 2 single 0.3333 0.3333 0.7003 units lattice
create_atoms 2 single 0.3333 0.3333 0.2997 units lattice

mass 1 97.9055000
mass 2 129.9067000

replicate 30 30 1

pair_style reax/c lmp_control checkqeq yes
pair_coeff * *ffield-MoTeCu Mo Te

timestep 0.5
thermo_style custom etotal pe ke temp evdwl press pxx pyy pzz pxy pxz pyz lx ly lz
thermo_modify line multi flush yes
thermo 10

variable ti equal 300.0

dump 1 all custom 100 traj.dump id type x y z q

fix 0 all qeq/reax 1 0.0 10.0 1.0e-6 reax/c
minimize 1.0e-8 1.0e-8 100 1000

velocity all create ${ti} 4928459 mom yes rot yes dist gaussian
fix 1 all nvt/sllod temp ${ti} ${ti} 100.0
fix 2 all deform 1 x erate 0.000002 remap v units box
run 100000
```



Nanomaterial mechanics explorer

1 Input → 2 Simulate

Mechanics Lab

Experiment: Melting

Melting

Melting Material: Nickel Nanoparticle

Crack Propagation

Phase Transition

Nanowire Tensile Test

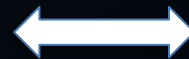
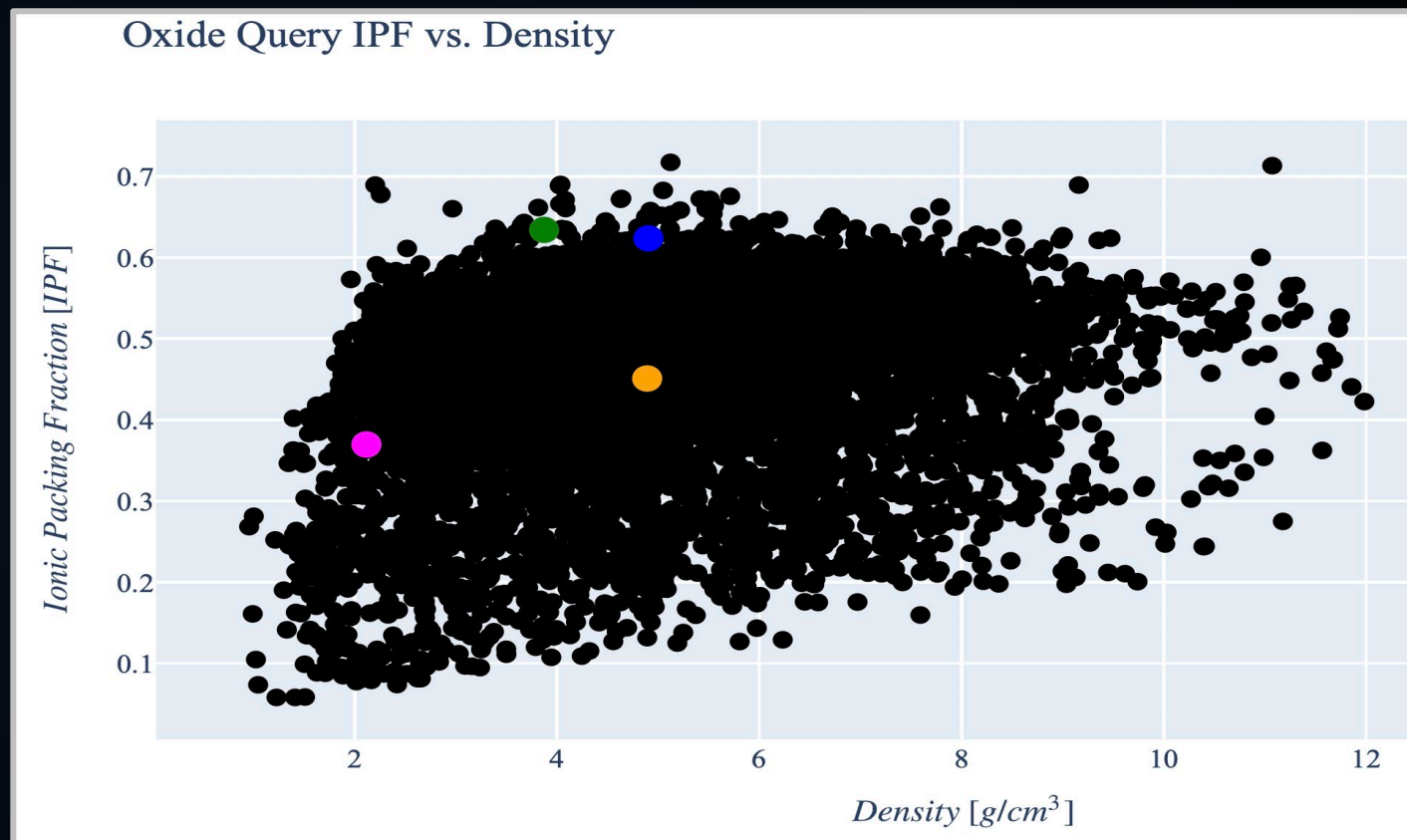
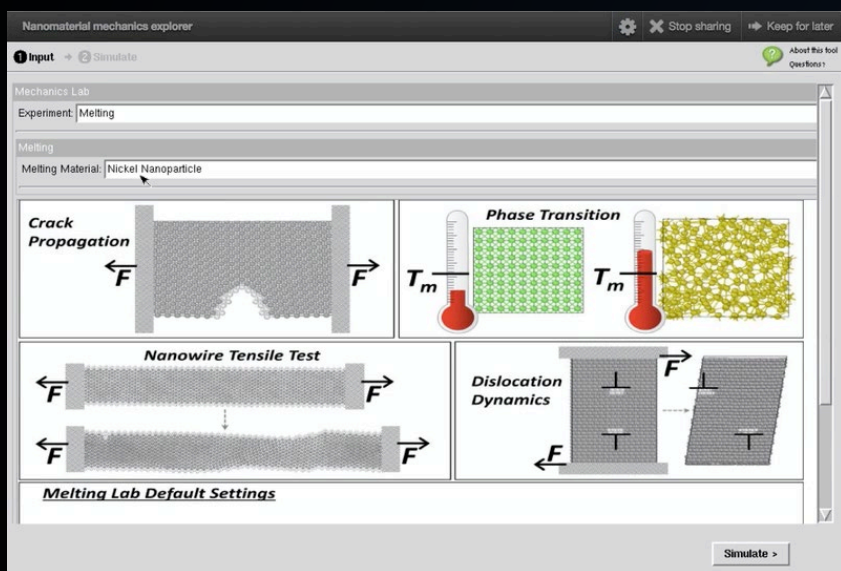
Dislocation Dynamics

Melting Lab Default Settings

Simulate >

Containerized Jupyter notebooks for ML

Compute and visualize properties for 65,000 oxides



The Materials Project

nanoHUB: a community platform

800+ apps & tools

170+ courses, 6500+ resources

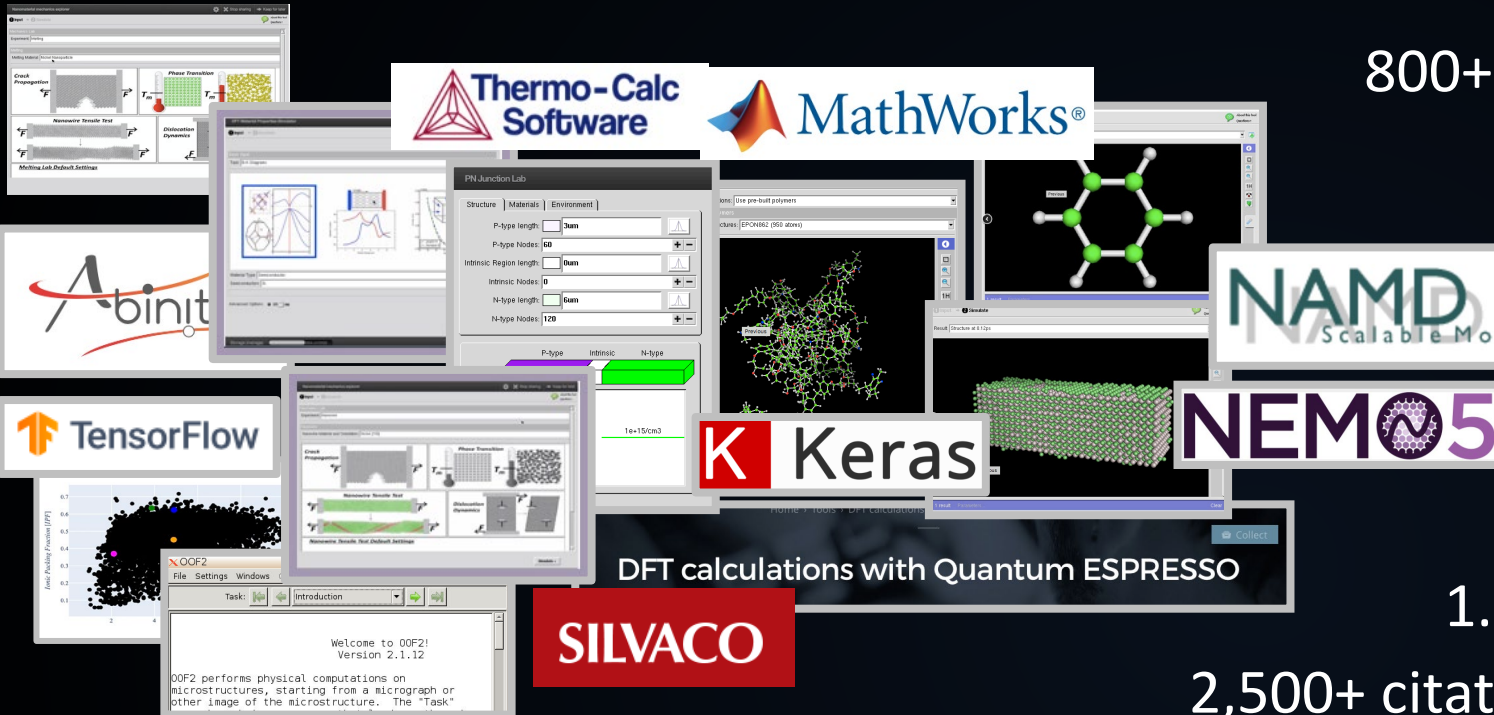
1,800+ contributors

250,000+ simulation users

1,000,000+ simulations per year

1.5 million visitors per year

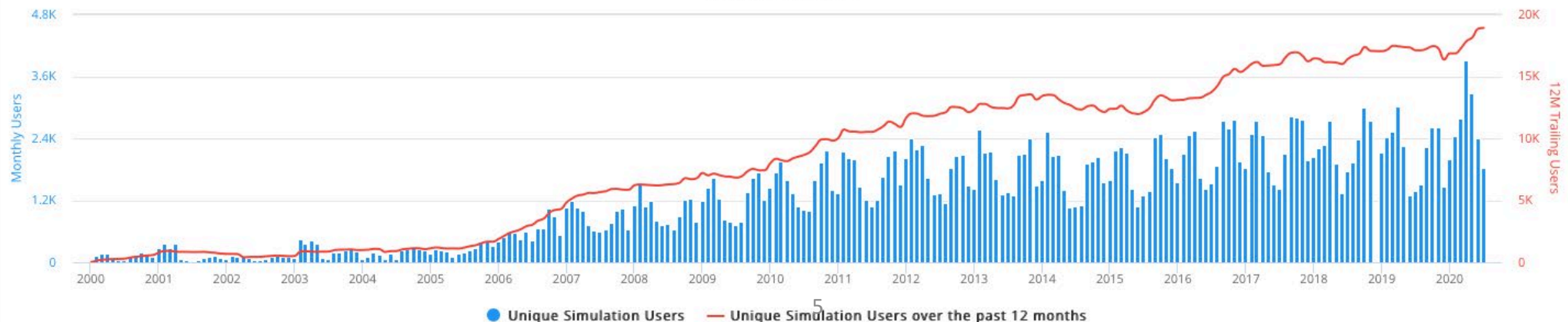
2,500+ citations



www.nanohub.org/usage

nanoHUB Simulation Users

12M Trailing and Monthly



We have a data problem

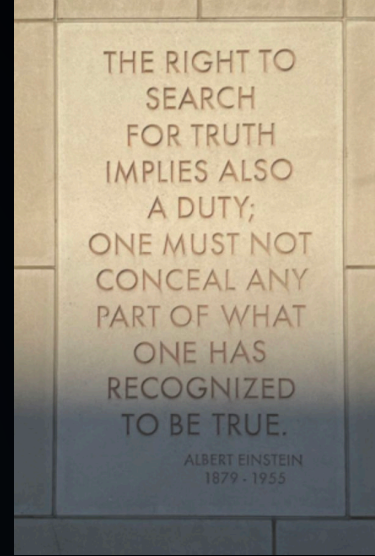
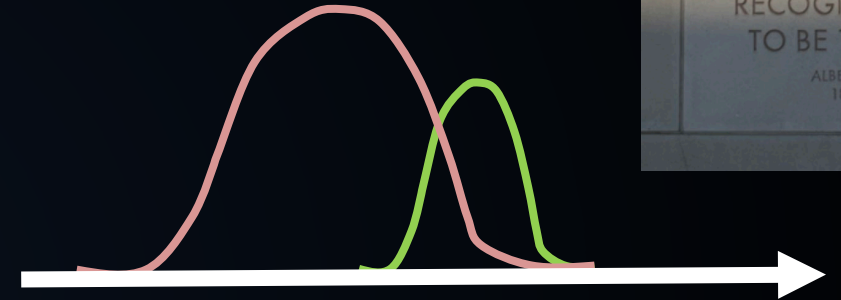
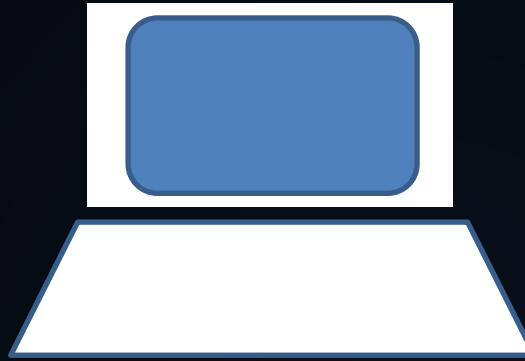
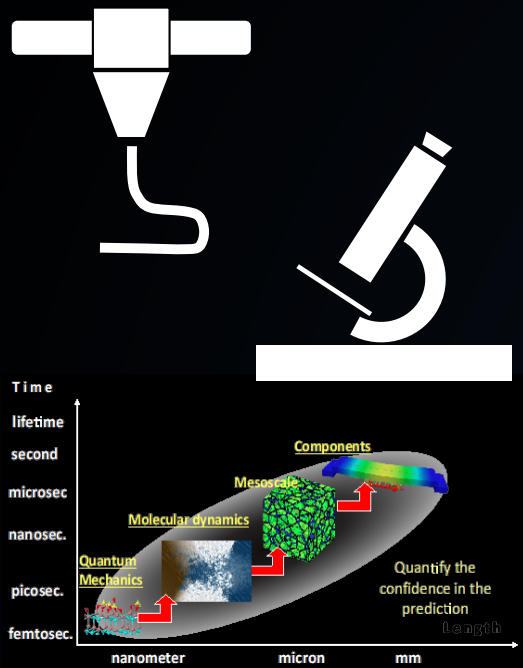
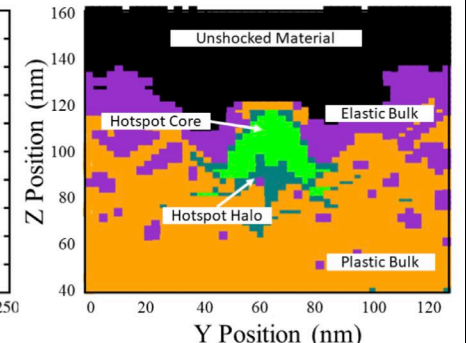
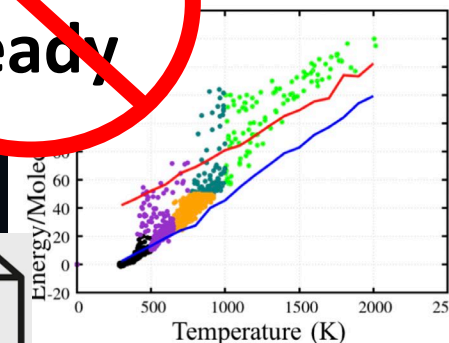


Table 1. Final Product Amounts for LLM-105 Using ReaxFF, Compared to Values for TATB and HMX^{93,94}

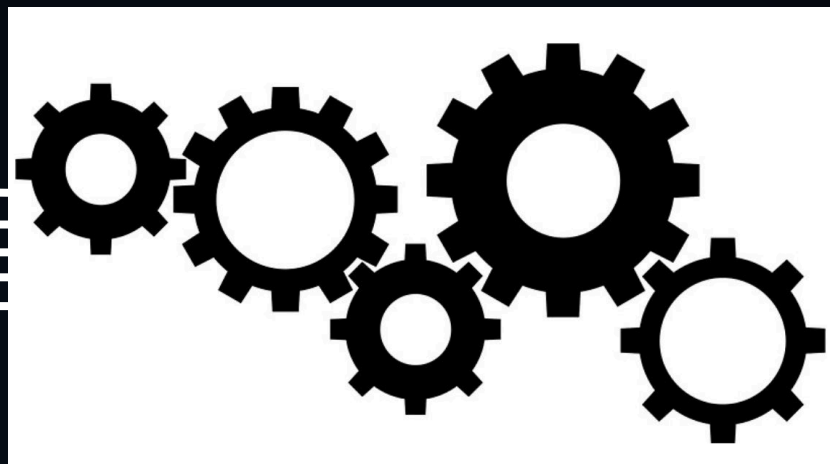
molecule	ReaxFF-2018 (mol/cm ³)	ReaxFF-LG (mol/cm ³)	LLM-105 Cheetah (mol/cm ³)	ReaxFF-LG TATB (mol/cm ³)	TATB Cheetah (mol/cm ³)	TATB exp. (mol/cm ³)	HMX Cheetah (mol/cm ³)	HMX exp. (mol/cm ³)
N ₂	1.109×10^{-2}	2.145×10^{-2}	2.604×10^{-2}	1.66×10^{-2}	2.17×10^{-2}	1.75×10^{-2}	2.47×10^{-2}	2.37×10^{-2}
CO ₂	1.313×10^{-2}	5.280×10^{-3}	1.297×10^{-2}	2.80×10^{-3}	1.20×10^{-2}	1.46×10^{-2}	1.24×10^{-2}	1.24×10^{-2}
H ₂ O	3.335×10^{-4}	7.442×10^{-3}	1.440×10^{-2}	1.05×10^{-2}	1.89×10^{-2}	1.56×10^{-2}	1.88×10^{-2}	2.05
NH ₃	2.264×10^{-3}	1.783×10^{-3}	1.194×10^{-3}	3.07×10^{-3}	1.64×10^{-3}	8.23×10^{-4}	1.97×10^{-3}	2.55



FAIR research/data workflows

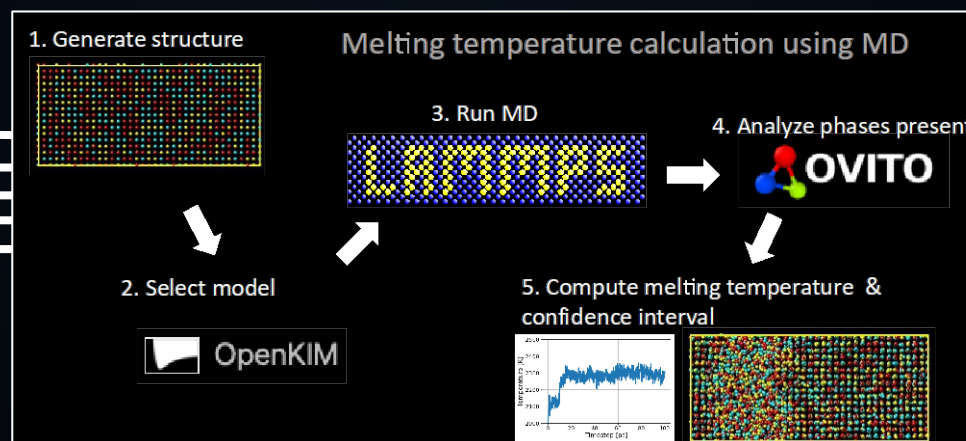


Raw data
Exp. conditions



Outputs

Inputs



Outputs

nanoHUB publishes your workflow & makes it FAIR

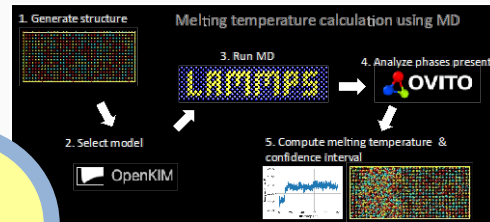
High Entropy Alloy Melting Point Calculation



By [Zachary D McClure¹](#); [Saaketh Desai¹](#); [Alejandro Strachan¹](#)

¹. [Purdue University](#)

Calculate melting point of high entropy alloys through phase coexistence method



Launch Tool

Version **1.7a** - published on 15 Aug 2023

doi:10.21981/QBT8-7M93 [cite this](#)

[Open source: license](#) | [download](#)

[View All Support](#)



Usage Citations Questions Reviews Wishlist Versions

Query &
discovery

World usage

Location of all "High Entropy Alloy
Melting Point Calculation" Users
Since Its Posting



nanoHUB publishes your workflow & makes it FAIR

+ reproducible

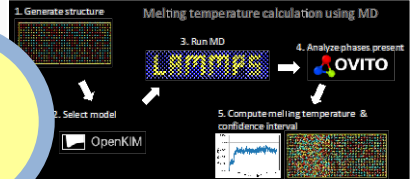
High Entropy Alloy Melting Point Calculation

By [Zachary D McClure¹](#); [Saaketh Desai¹](#); [Alejandro Strachan¹](#)
1. [Purdue University](#)

Calculate melting point of high entropy alloys through phase coexistence method

Launch Tool

Version **1.7a** - published on 15 Aug 2023
doi:10.21981/QBT8-7M93 cite this
Open source: [license](#) | [download](#)
[View All Support](#)



World usage

Location of all "High Entropy Alloy Melting Point Calculation" Users Since Its Posting



Map Satellite

Google

Keyboard shortcuts | Map data ©2024 | Terms

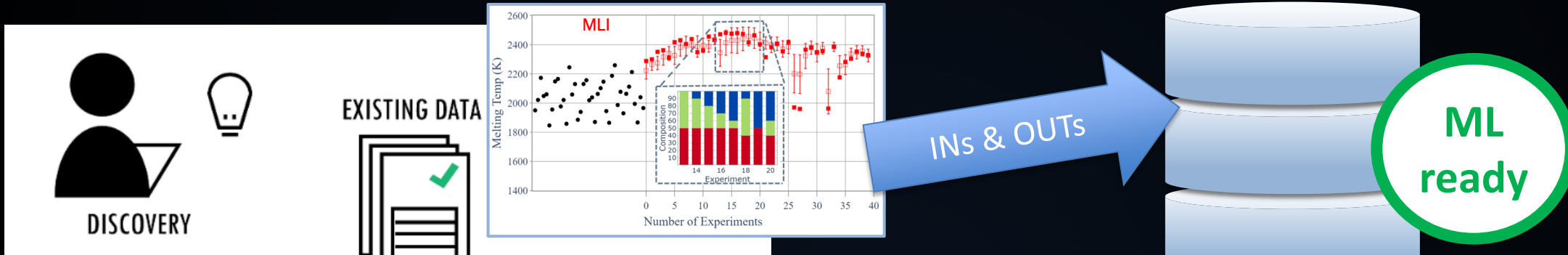
Query &
discovery

Containerized

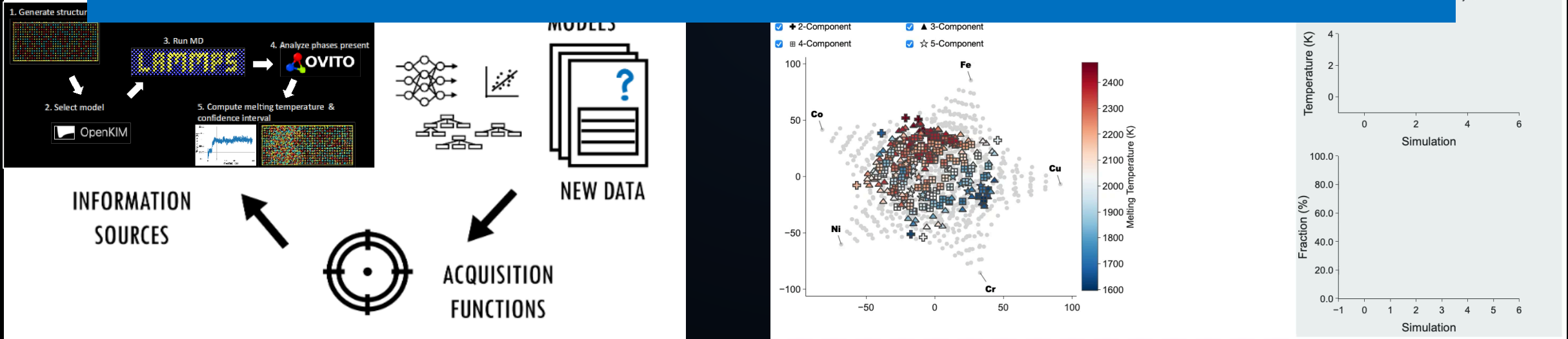
Simulation as a service and automatic FAIR data

Autonomous ML workflow

ResultsDB



FAIR data & workflows accelerate innovation



Experimental data analysis workflow

Mishra et al. Comp Mat Sci 233, 112671 (2024)

Collect & upload exp. data & metadata



Automatic model calibration & selection
(42 models)



EXPLORE PUBLISH COMMUNITY ABOUT SUPPORT

Refractory Oxidation Database

By Saswat Mishra, Sharmila Karumuri, Michael S Titus, Ilias Bilonis, Alejandro Strachan

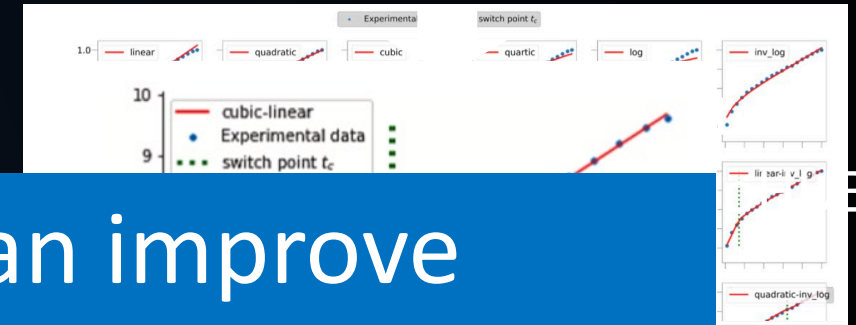
Creates a database of refractory alloys oxidation

Launch Tool

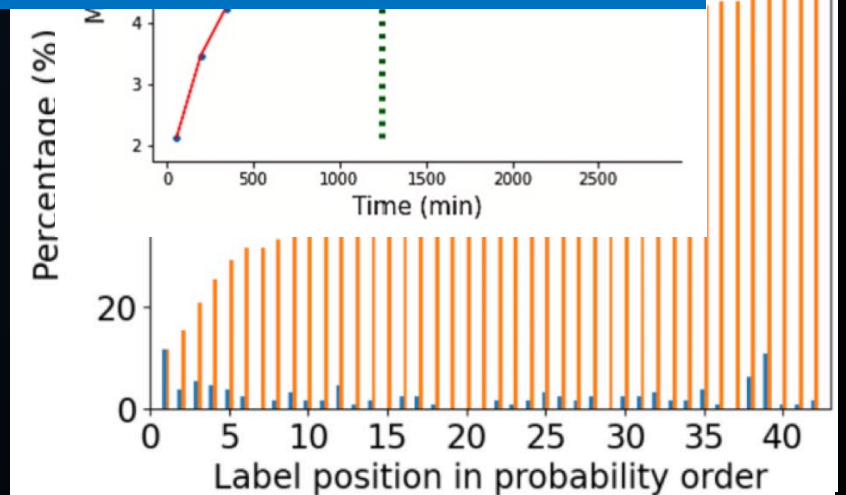
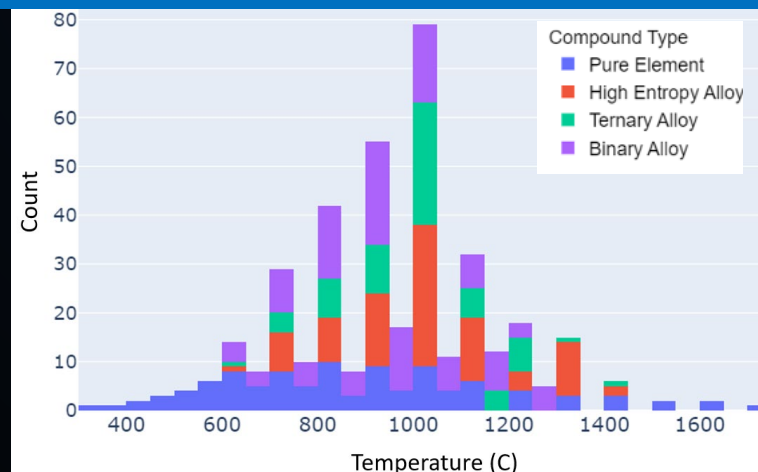
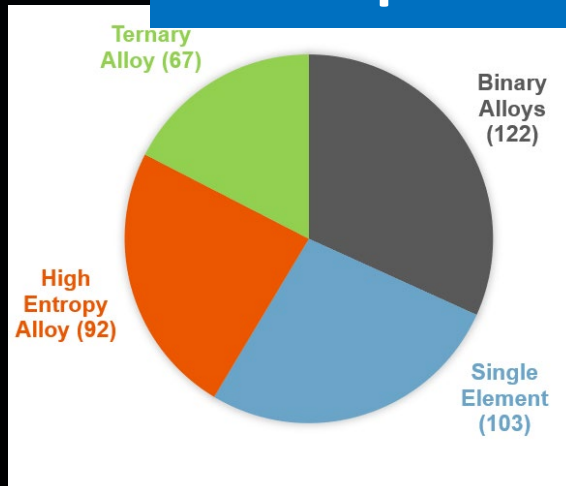
Version 2.2 - published on 14 Nov 2022

doi:10.21981/1H5E-7J81 cite this

Open source: license | download



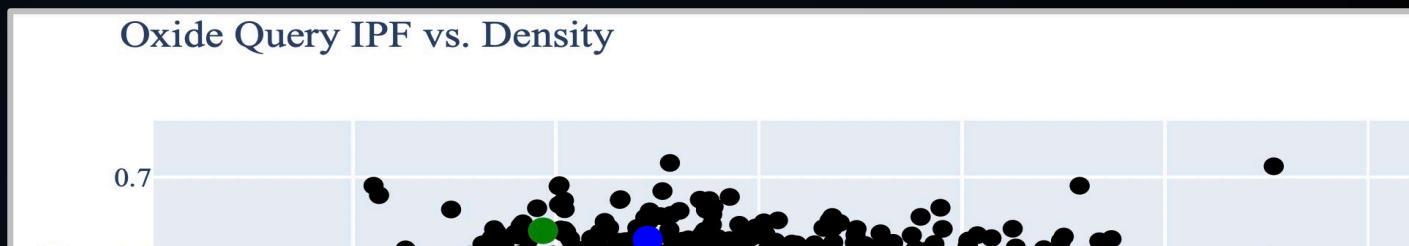
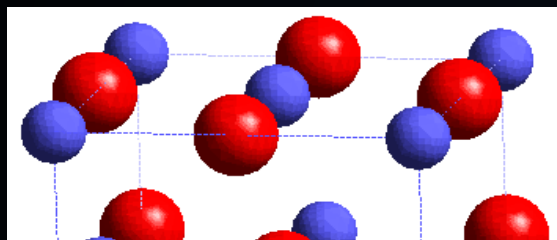
Data analysis workflows can improve
reproducibility and consistency across labs



B's

Harnessing data resources for education

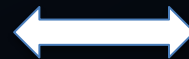
Compute and visualize properties for 65,000 oxides



Enhance education & workforce development
Hands-on, active learning
Provide our citizens with the knowledge to benefit
from open data

$$IPF = \frac{V}{(a \times b) \cdot c}$$

Density [g/cm³]



The Materials Project

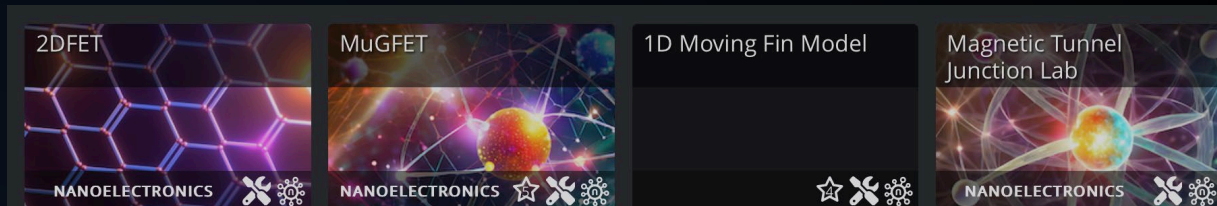
User: *I am a materials scientist interested in machine learning. Which tools should I try?*

nanoCHAT: *Based on your interest in machine learning as a materials scientist, I recommend trying the following tools:*

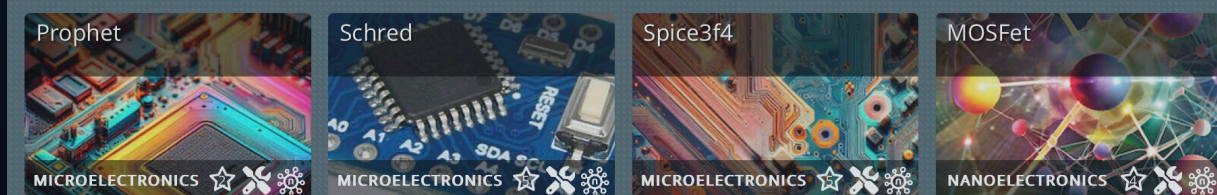
- 1. Machine Learning for Materials Science: Part 1 (mseml) - This tool offers practical examples and tutorials on machine learning and data science for materials science. It's a great starting point for learning the basics and exploring the applications of machine learning in materials science.*
- 2. Materials Simulation Toolkit for Machine Learning (MAST-ML) tutorial - This tutorial provides a comprehensive introduction to using machine learning methods in materials science. It covers various aspects of machine learning, including data preparation, model selection, and uncertainty quantification.*
- 3. Machine Learning Lab Module (intromllab) - This lab activity provides an overview of a materials science machine learning workflow, from data cleaning to model analysis. It's a valuable resource for learning the practical aspects of machine learning in materials science.*
- 4. Machine Learning Force Field for Materials (maml) - This tool is designed for developing Machine Learning Interatomic Potentials (ML -*

Finding stuff

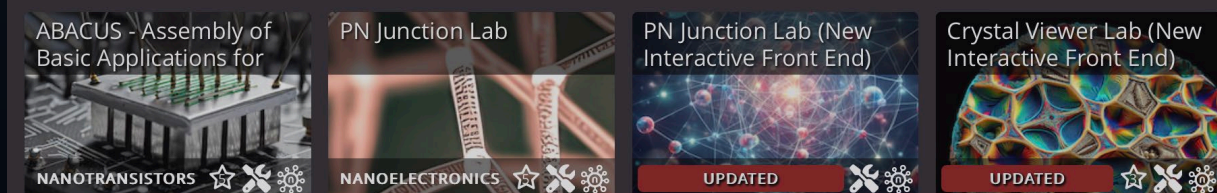
Personalized recommendation engine



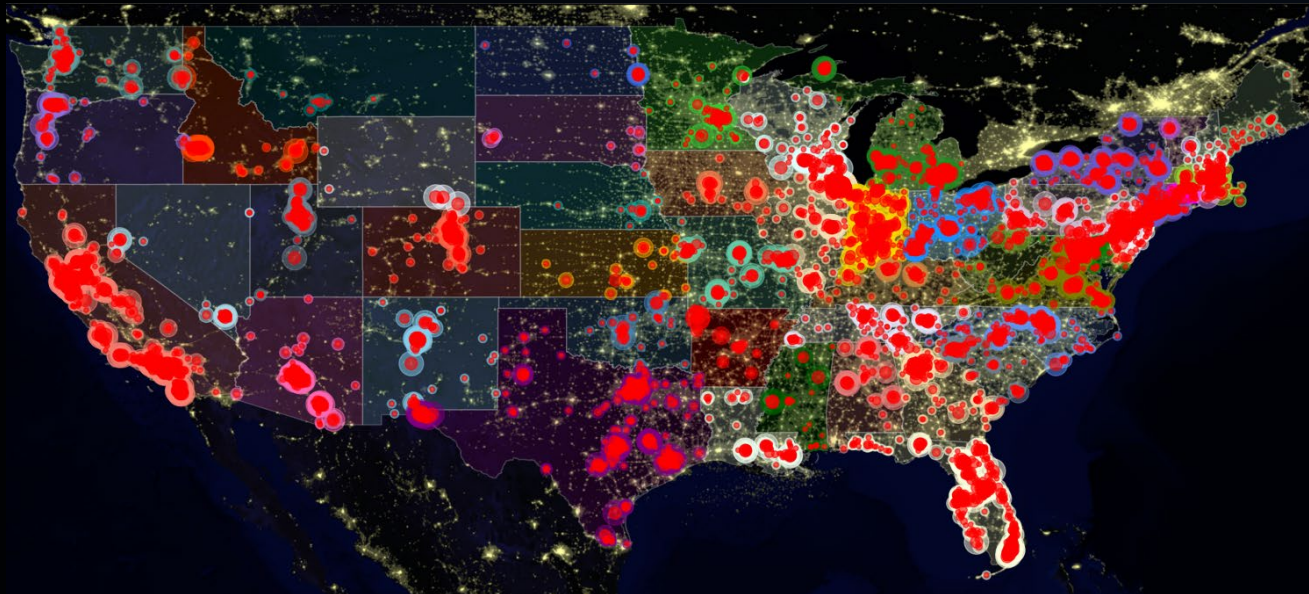
USERS LIKE YOU ENJOY



MOST POPULAR



Impact



- 250+ simulation users served
- nanoHUB resources used in 77% of technical minority-serving institutions

Organization	Sim Users
Purdue University Main Campus	19,963
Arizona State University Tempe	7,812
University of Illinois at Urbana Champaign	4,984
Northwestern University	4,511
Massachusetts Institute of Technology	2,341
Florida Agricultural and Mechanical University	1,908
Harvard University	1,653
Stanford University	1,544
Cornell University	1,445
University of California Berkeley	1,439
Ball state university	1,346
The University of Texas at El Paso	1,284
University of California Los Angeles	1,103
North Carolina State University at Raleigh	1,023
Chicago State University	999
University of Michigan Ann Arbor	965
...	
Total	91,955

The Materials Research Data Alliance



MDI Provider Integration and Interoperability

WG1

The mission of the MDI Provider Integration and Interoperability WG is to promote technical awareness about U.S. and international projects specifically among MDI developers.

[LEARN MORE](#)



Documenting Interoperable Data and Modeling Resources

WG2

The last decade saw an explosion of cyber-infrastructure for materials data and modeling in academia, national labs, and the private sector.

[LEARN MORE](#)



Workflow Interoperability

WG3

Materials researchers are often encouraged or required to make data available and FAIR.

[LEARN MORE](#)



WG4: Materials Data Repository Priorities (pending)

WG4

Repositories of materials data are fundamental to MGI efforts.

[LEARN MORE](#)



Data Dictionaries Working Group

WGS

Data dictionaries are central to making metadata transferable and context relationships explicit.

[LEARN MORE](#)



WG 6: FAIR for Materials

WG6

This group will describe the current state of materials data efforts around the world, discuss what FAIR data principles mean in the context of materials science and engineering, summarize inputs garnered from conversations with key community stakeholders, and propose a roadmap with concrete actions, both individual and collective, towards FAIR materials data.

[LEARN MORE](#)

NSF FAIROS program

- FAIR Data
- FAIR Models
- FAIR Train
- FAIR Train



MRS Bulletin

Impact Opinion & Perspective
MRS Bulletin(2023): 1-5.

**Community action on FAIR data
will fuel a revolution in materials
research**



L. Catherine Brinson,*^{ID} Laura M. Bartolo, Ben Blaiszik, David Elbert,
Ian Foster, Alejandro Strachan, and Peter W. Voorhees

2024 MARDa VIRTUAL ANNUAL MEETING
FEBRUARY 20-22

Meeting Chairs: Dr. Debra Audus (NIST), Dr. Corey Oses (Johns Hopkins), Dr. Fatih Sen (Noverais)

The 2024 Annual Meeting provides a deep dive into the pivotal role of open data in shaping the landscape of AI/ML applications and their future trajectories. Join us for engaging discussions on collaborative, interagency initiatives aimed at advancing FAIR data principles and the expansive integration of AI in materials discovery, design, and deployment. The meeting includes ~4 hours of short presentations, panel discussions, and workshops each day. Participants are welcome to attend any or all days. Registration and membership in MaRDA is free and open to all materials data stakeholders interested in the mission of MaRDA.

Registration now open! [Registration Link](#) (all participants should be a [MaRDA member](#) prior to event registration)

Final thoughts

- Data and data workflows should be made FAIR (and reproducible)
- Making data FAIR cannot be a manual afterthought
- Include positive and negative results
- Agencies should do a better job at evaluating and enforcing data mandates
- Community needs to embrace re-use and fight the urge to re-invent the wheel
 - This can be the key to sustainability
- Modern data practices should be incorporated/used in undergraduate education