

nanoHUB: getting started guide to tool developers

Develop and publish tools in nanoHUB

Make your research reproducible and your workflows and data FAIR

Tanya Faltens, Daniel Mejia, Steven Clark, Juan Carlos Verduzco & Ale Strachan*

* strachan@purdue.edu

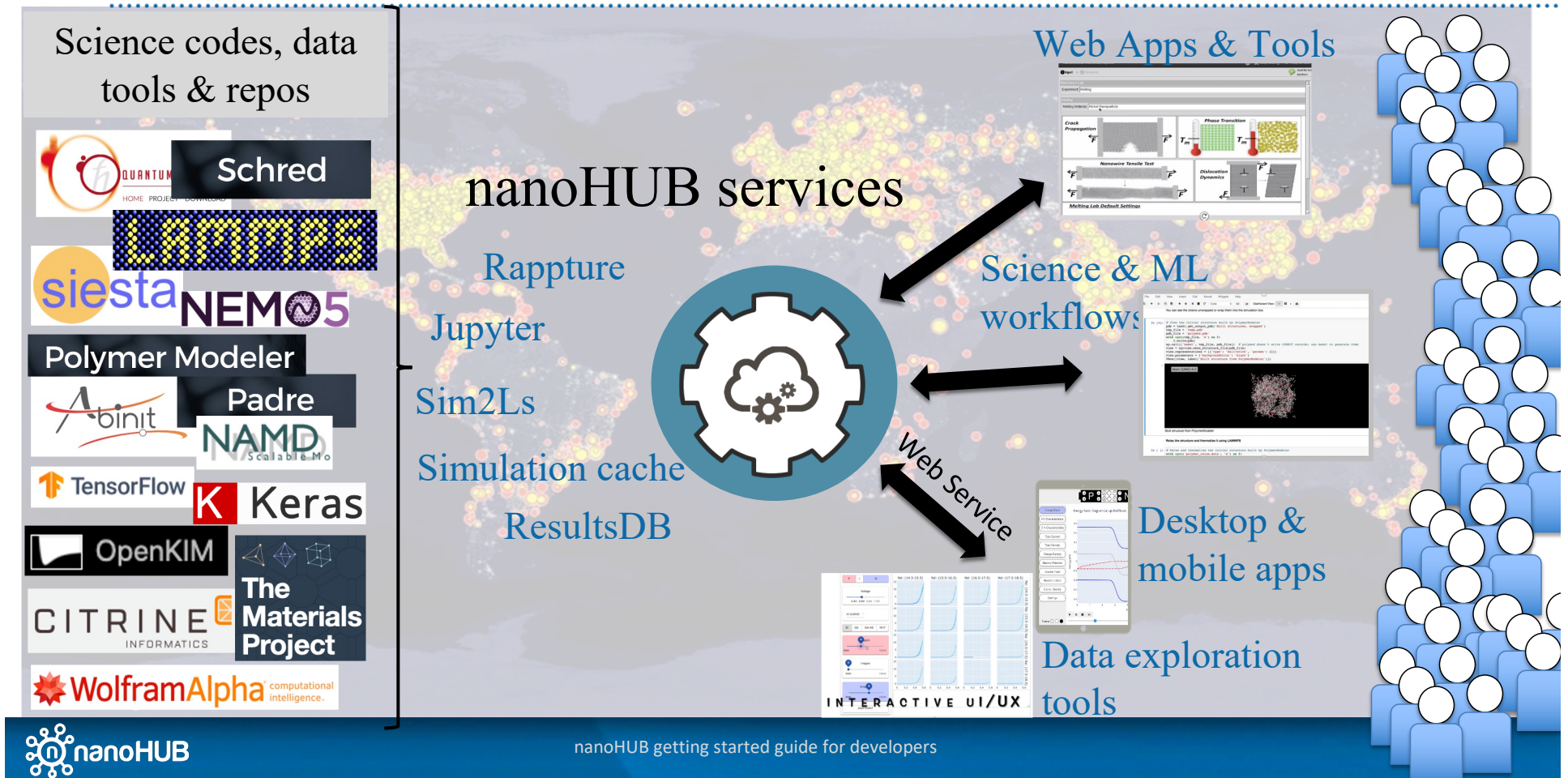
School of Materials Engineering &

Purdue University

West Lafayette, Indiana USA



nanoHUB: online apps, tools, & data



Overview



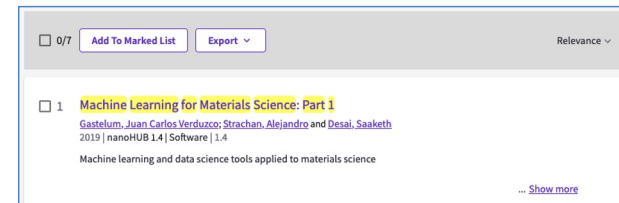
1. Why publish tools & apps in nanoHUB?
 - Tools are publications (DOIs and indexed by Web of Science)
 - Share your work with your community (22,000+ annual sim users)
2. Various tool and app types
 - Apps, workflows, Jupyter notebooks, commercial codes, X11 GUIs
3. Sim2Ls, FAIR workflows and data
 - Develop and publish Sim2Ls
4. Developing Apps
 - Connecting Sim2Ls to Jupyter and Web Apps
5. Tool Publication process
 - Register, deploy, test, and publish
6. Development environment
 - A Unix development environment (Jupyter or Linux desktop)
7. Simulation and data as a service
 - Launching tools and querying the ResultsDB

Why publish in nanoHUB?

Share your code/workflow with your community & make it:

1. Findable

- Tools are publications and are assigned a digital object identifier (DOI) & indexed by Web of Science



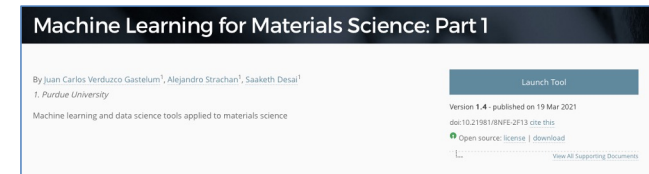
2. Accessible

- Online using a web browser
- Using standard web services



2. Interoperable

- Publish complete workflows & use metadata to make services and requirements discoverable



4. Reusable

- Full workflows and data are open and accessible

5. Reproducible

- Apps, tools, & workflows are containerized with their compute environment

Why publish in nanoHUB?

Share your code/workflow with your community & make it:

1. Findable

- Tools are publications and are assigned a digital object identifier (DOI) & indexed by Web of Science

2. Accessible

- Online using a web browser
- Using standard web services

2. Interoperable

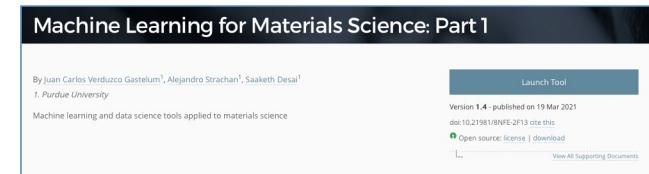
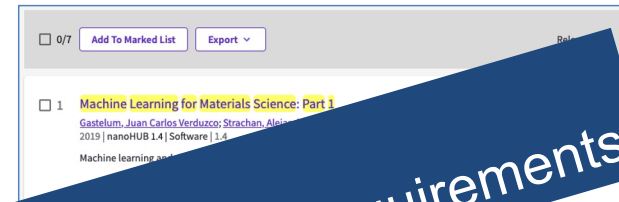
- **Publicly available metadata to** requirements discoverable

4. Reusable

- Full workflows and data are open and accessible

5. Reproducible

- Apps, tools, & workflows are containerized with their compute environment



Why publish in nanoHUB? (cont'd)

4. Track your impact with detailed user metrics

Machine Learning for Materials Science: Part 1

By Juan Carlos Verdusco Gastelum¹, Alejandro Strachan¹, Saaketh Desai¹
¹ Purdue University

Machine learning and data science tools applied to materials science

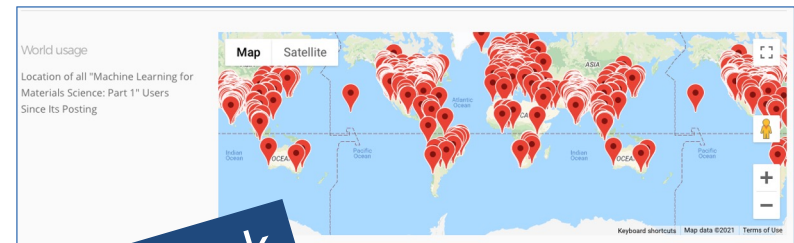
Version 1.4 - published on 19 Mar 2021
doi:10.21961/ANFE-2F13 cite this
Open source license | download
View All Supporting Documents

2,561 users, detailed usage
0 Citations(s)
1 question (Ask a question)
3 reviews(s) (Review this)
0 wishes(s) (New Wish)
Share

Materials Science

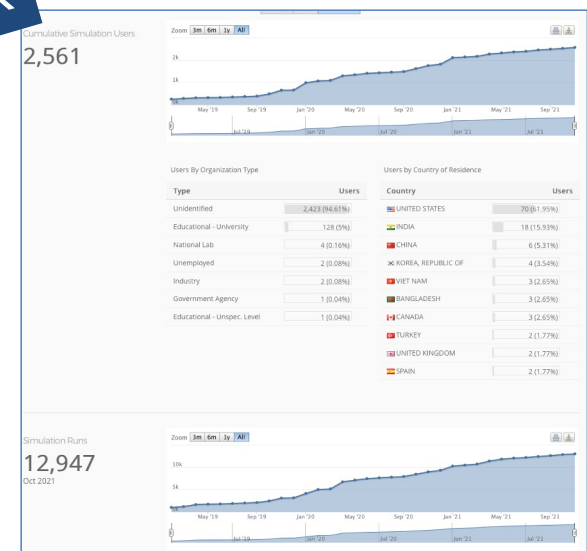
Category: Tools
Published on: 19 Mar 2021

Abstract
Data science and machine learning are playing increasingly important roles in science and engineering, and machine learning is becoming a key tool for materials scientists. This online tool provides machine learning examples in the field of materials science using Jupyter notebooks, which are interactive documents that contain code and text. The initial set of tutorials focus on:
i) data query, organization and visualization
ii) developing a simple model using linear regression to explore correlations between materials properties
iii) neural network models trained to predict materials properties from basic element properties
Suggested exercises are included in each Jupyter notebook.
This tool was used in the Hands-on Machine Learning and Data Science Training Workshop conducted by nanoHUB in April 2020. Offerings for the tutorial can be found in nanoHUB resources here, here and here.



Assess the impact of your work

<https://nanohub.org/tools/mseml>



5. For certain tools outputs and results are automatically saved

- **Simulation cache** provides fast results for previously run simulations
- **Results database** enables querying all prior results