Developing and deploying Jupyter tools in nanoHUB

In this workshop

- Develop Jupyter notebooks using nanoHUB
 - Deploying tools in nanoHUB

Alejandro Strachan

strachan@purdue.edu

School of Materials Engineering & Network for Computational Nanotechnology Purdue University

West Lafayette, Indiana USA

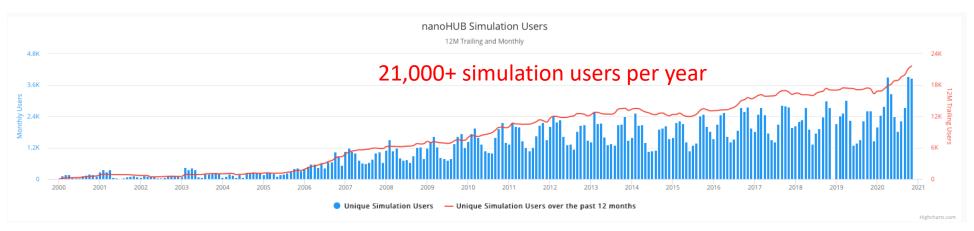




nanoHUB

Why deploy tools in nanoHUB?

- Tools are publications: indexed by Web of Science and Google Scholar.
- Impact statistics: usage metrics, secondary citations
- Anyone can run your tool without downloading/installing any software
 - Use it in the classroom
 - Document your research and make it reproducible (tools are containerized)
- A full scientific software development environment
 - Libraries for machine learning, statistics, data science, visualization, etc.
- Launch physics-based, research grade codes
 - LAMMPS, Quantum Espresso, NEMO5, abinit, MOOSE, and many more



What is Jupyter?

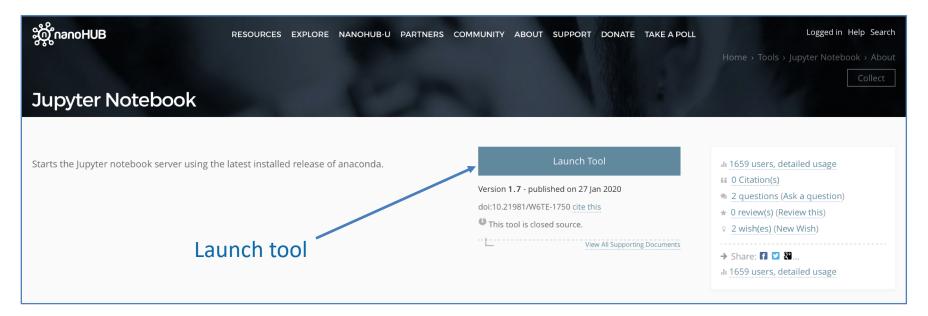
Popular tool for interactive computing

Rich text The standard deviation can be easily calculated: $\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$ In our case, $\sigma = (2/3 \times 0.5^3)^{1/2} \sim 0.288$ In the following cells we will study how large a sample we need to obtain accurate estimates of the mean and standard deviation of the population. In [20]: #Create empty arrays to collect our data SampleSize = []
SampleMean = [] In [17]: # Let's start by plotting a histogram of the sample v SampleStd = []# Loop over increasingly larger sample sizes, compute SAMPLE mean and standard do # The matplotlib command below creates and plots a hi for size in [25, 50, 75, 100, 150, 500, 1000, 1500, 2000, 5000]: # The input is the number of bins in the histogram sample = sample = np.random.uniform(0,1,size) plt.hist(sample,30) SampleSize = np.append(SampleSize,[size]) plt.ylabel("Counts") plt.xlabel("Value") SampleMean = np.append(SampleMean,[np.mean(sample)]) Live code SampleStd = np.append(SampleStd,[np.std(sample,ddof=1)]) plt.show() # The histogram you obtain should be relatively consi # Fluctuations are expected, if you increase the size <u>بر</u> 25 ලි 20 **Powerful** 10 visualization



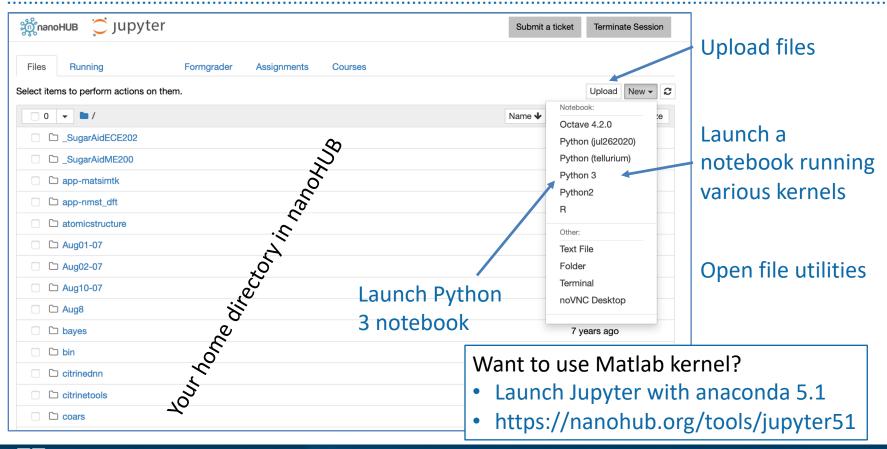
Start working with Jupyter on nanoHUB

https://nanohub.org/tools/jupyter (you will need a free nanoHUB account to launch Jupyter)

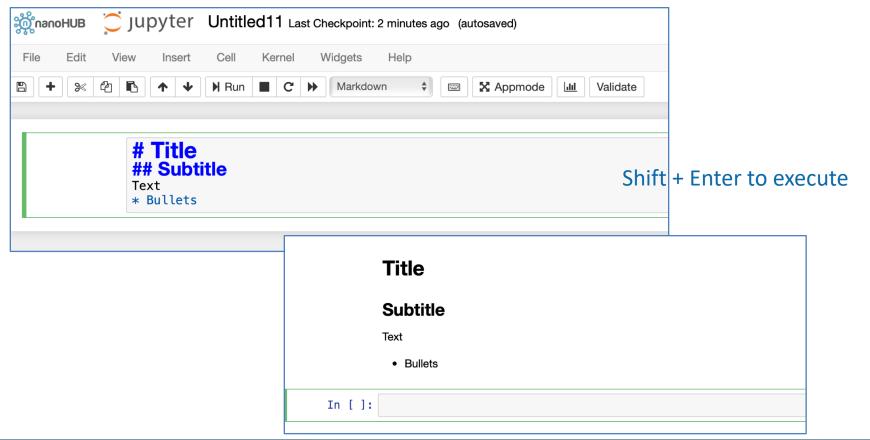




Inside the Jupyter tool



Markdown cells

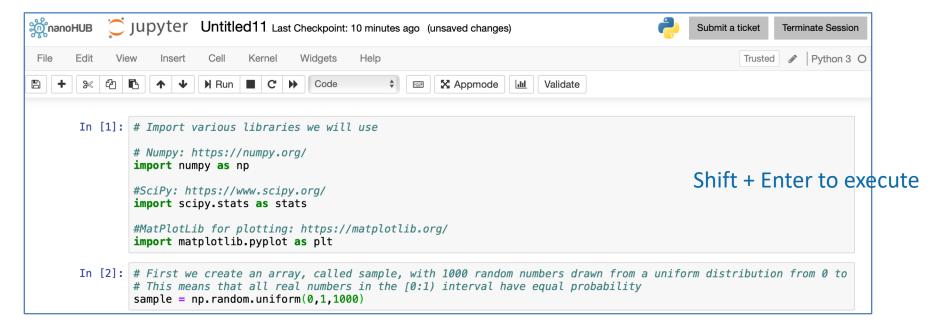




Code cells

Import libraries we will use

- Lots of libraries available
- Submit a Ticket if we don't have what you need



Generate an array of random numbers



Plotting

Now let's make a plot In [3]: # Let's start by plotting a histogram of the sample we just obtained. # The matplotlib command below creates and plots a histogram with our sample. # The input is the number of bins in the histogram plt.hist(sample,30) plt.ylabel("Counts") plt.xlabel("Value") plt.show() 50 40 20 10 0.0 0.2 0.4 0.6 0.8 1.0 Value

Shift + Enter to execute



Math in markdown

```
The *population* average, \infty, and standard deviation, \simeq, are defined as: \infty = {\{\sum_{i=1}^{N}\} x_i \in \mathbb{N}\} } (This is the same as that of the sample, but the sum runs over all members of the population.) \simeq \sigma^2 = {\sum_{i=1}^{N}} (x_i - \sum_{i=1}^{N}) (x_i - \sum_{i=1}^{N}) Note the \simeqN$ in the denominator.
```

Shift + Enter to execute

The *population* average, μ , and standard deviation, σ , are defined as:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

(This is the same as that of the sample, but the sum runs over all members of the population.)

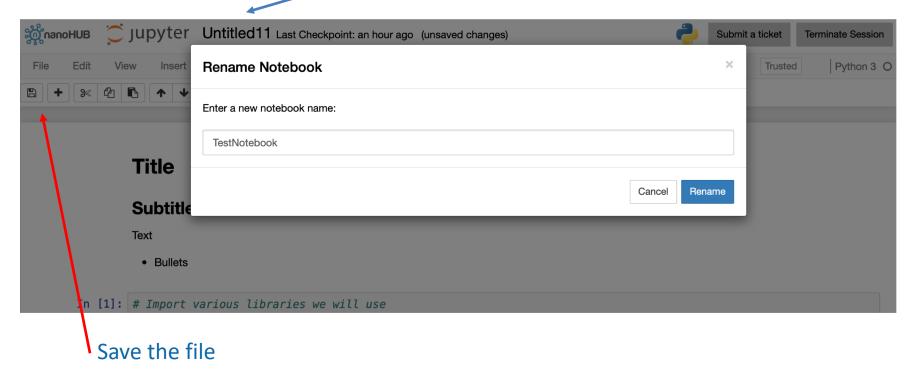
$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N}$$

Note the N in the denominator.



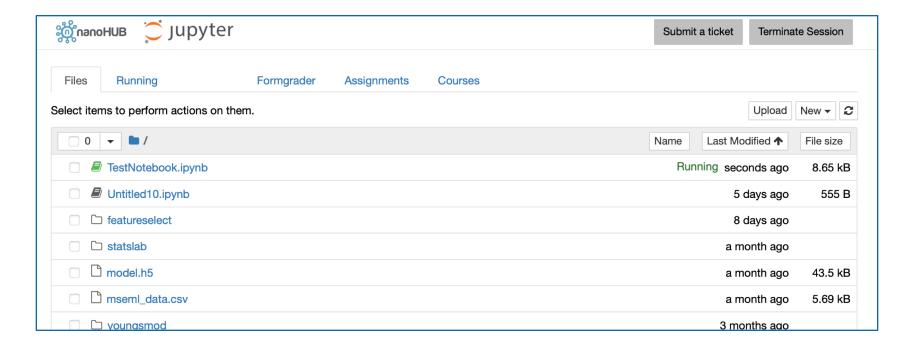
Saving your notebook before publishing

Click on the name to rename file





Check your home directory



Deploying a tool in nanoHUB

Tools are publications

By Juan Carlos Verduzco Gastelum¹, Alejandro Strachan¹, Saaketh Desai¹

1. Purdue University

Machine learning and data science tools applied to materials science

Version 1.3 - published on 01 Apr 2020
doi:10.21981/WGQC-3249 cite this
Open source: license | download

View All Supporting Documents

Al. 1783 users, detailed usage
66 O Citation(s)
9 1 question (Ask a question)
9 2 erview(s) (Review this)
9 0 wish(es) (New Wish)

Share: 1 2 8 ...

With DOIs

Researchers should cite this work as follows:

Juan Carlos Verduzco Gastelum, Alejandro Strachan, Saaketh Desai (2020), "Machine Learning for Materials Science: Part 1," https://nanohub.org/resources/mseml. (DOI: 10.21981/WGQC-3249).

BibTex EndNote

Indexed by Web of Science and Google Scholar



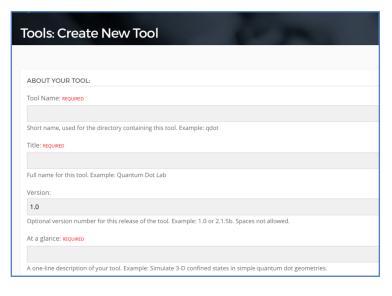


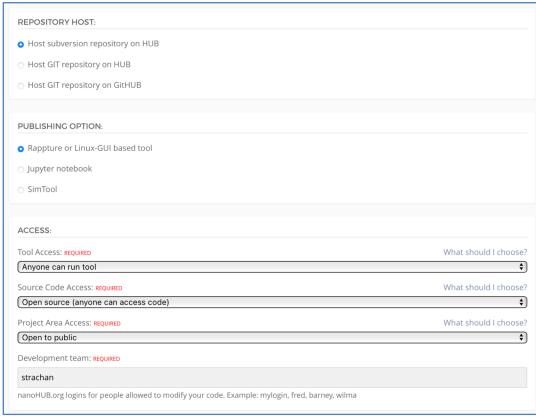


Ale {

Publish a tool

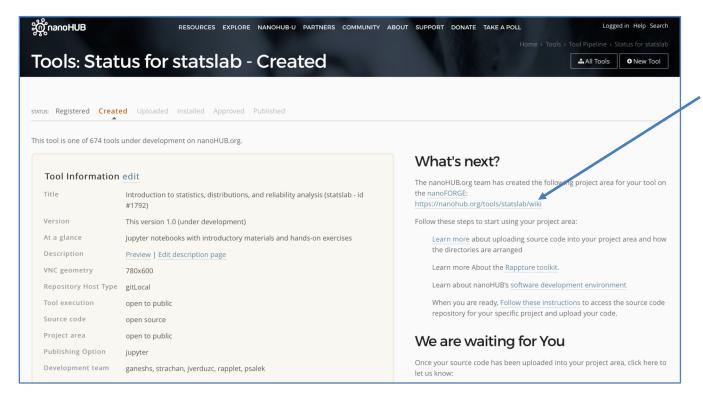
https://nanohub.org/tools/create







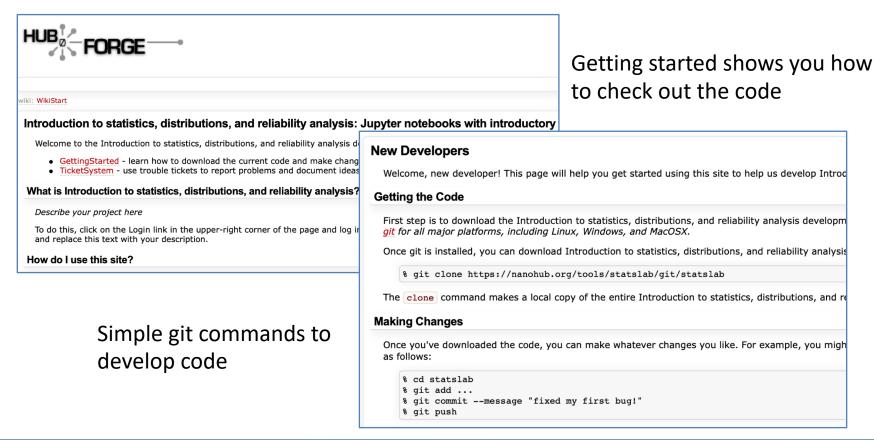
Tool registered by nanoHUB – ready for you to add code



Click link following nanoFORGE

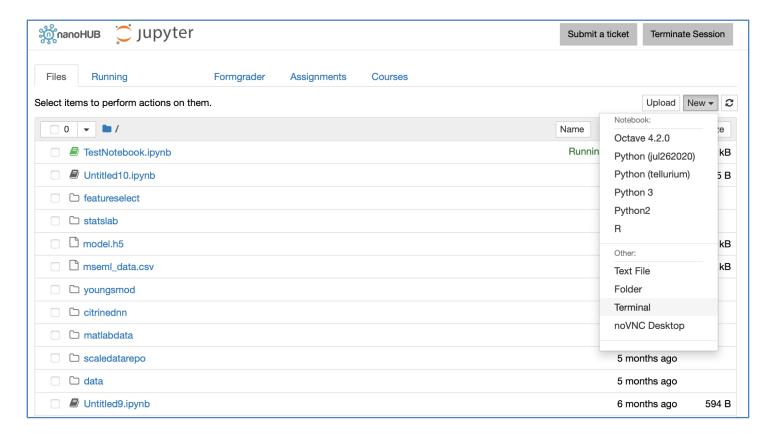


Tool registered by nanoHUB – ready for you to add code





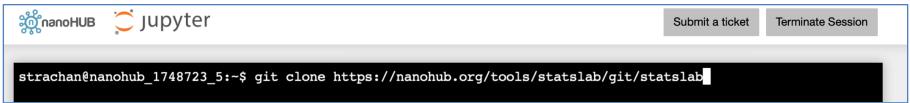
Open a terminal to move the code to the repository





Setup your tool

Check out (clone) the empty tool directory structure



Check out the directory structure (statslab should be change to your tool's short name)



Copy your notebook to the bin directory:

>cp../TestNotebook.ipynbbin/



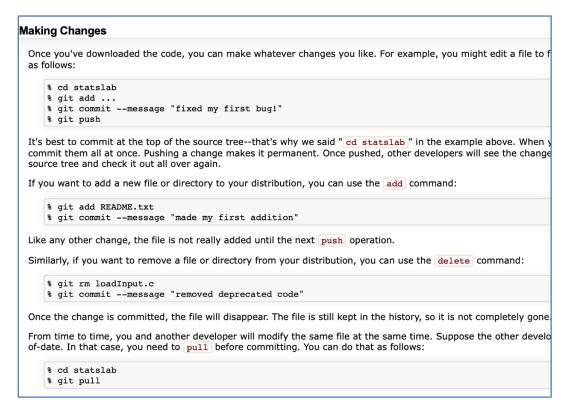
The bin & middleware directories

```
strachan@nanohub_1748723_5:~/statslab$ ls -ltr bin/
total 140
-rw-r--r-- 1 strachan public 14997 Dec 9 13:05 Module3.ipynb
-rw-r--r-- 1 strachan public 15314 Dec 9 13:07 LandingPage.ipynb
-rw-r--r-- 1 strachan public 60309 Dec 9 13:22 Module2.ipynb
-rw-r--r-- 1 strachan public 45365 Dec 14 21:09 Module1.ipynb
strachan@nanohub_1748723_5:~/statslab$
```

```
strachan@nanohub_1748723_5:~/statslab$ vi middleware/invoke strachan@nanohub_1748723_5:~/statslab$
```



When you are done, commit the code

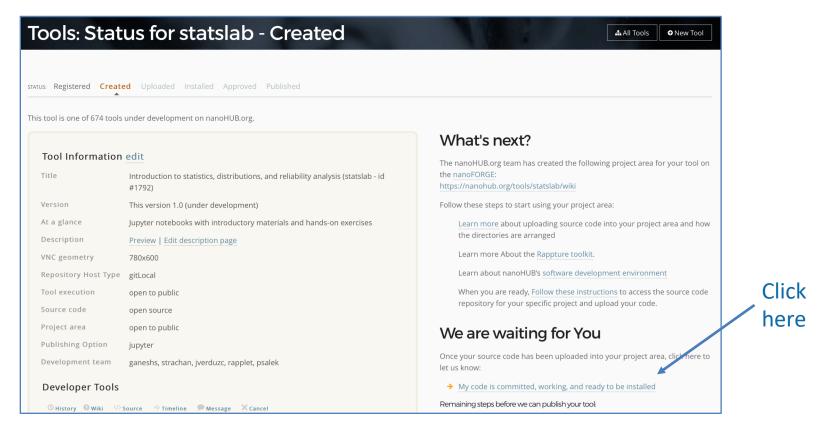


Add files Commit changes Push your changes

Use "pull" to get the most updated version of the code

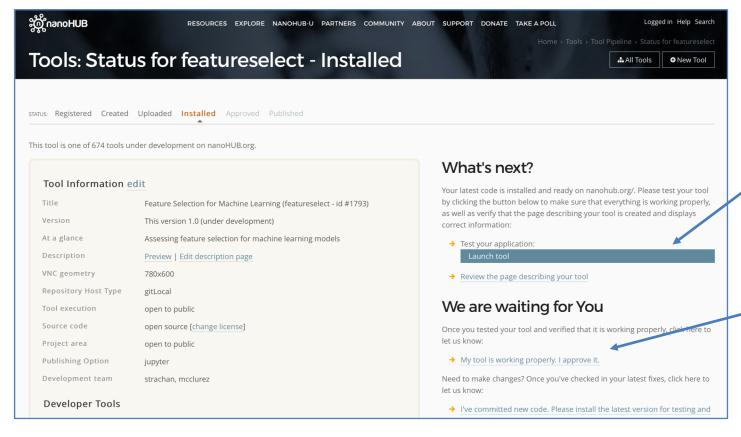


Let nanoHUB know when you are done





Try your tool before deploying it



Test your tool

When you are done, click here.

Done!



Check out what is available

Libraries available Jupyter

```
In [4]: import pkg_resources
        dists = [str(d) for d in pkg_resources.working_set]
       print(dists)
```

['-illow 5.4.1', 'zstandard 0.11.0', 'zipp 0.6.0', 'zict 0.1.4', 'yarl 1.3.0', 'yamlmagic 0.2.0', 'xmltodict 0. ['-illow 5.4.1', 'zstandard 0.11.0', 'zipp 0.6.0', 'zict 0.1.4', 'yarl 1.3.0', 'yamlnagic 0.2.0', 'xmltodict 0.12.0', 'xlwt 1.3.0', 'XlsxWriter 1.1.5', 'xlrd 1.2.0', 'xgboost 0.90', 'xarray 0.11.3', 'wurlitzer 1.0.2', 'wrapt 11.11', 'wolframalpha 3.0.1', 'widgetsnbextension 3.5.1', 'whichcraft 0.5.2', 'wheel 0.34.2', 'Werkzeug 0.1 4.1', 'websocket-client 0.56.0', 'webencodings 0.5.1', 'wcwidth 0.1.7', 'watchdog 0.9.0', 'w3lib 1.20.0', 'vtk 8.1.2', 'virtualenv 16.5.0', 'vine 1.3.0', 'urllib3 1.25.9', 'unicodecs v 0.14.1', 'uncertainties 3.0.3', 'typed -ast 1.4.0', 'twine 1.13.0', 'traittypes 0.2.1', 'traitlets 4.3.2', 'tqdm 4.47.0', 'tox 3.9.0', 'tornado 5.1.1', 'toolz 0.9.0', 'toml 0.10.0', 'tinydb 3.15.2', 'threadpoolctl 2.1.0', 'tf-estimator-nightly 1.14.0.dev2019030 115', 'textwrap3 0.9.2', 'testpath 0.4.2', 'tesedml 0.4.5.0', 'tesbml 5.18.0', 'terminado 0.8.1', 'termicolo 0.8.1', 'termicolo 0.8.1', 'termicolo 0.8.1', 'termicolo 1.1.0', 'tenuml 1.1.1.2', 'tensorflow 1.13.1', 'tensorflow-estimator 1.13.0', 'tensorboard 1.13.1', 'tenacity 5.0 4', 'tellurium 2.0.3', 'tecombine 0.2.3.0', 'tblib 1.3.2', 'tb-nightly 1.14.0a20190301', 'tabulate 0.8.3', 'tables 3.5.1', 'sympy 1.3', 'statsmodels 0.9.0', 'sshtunnel 0.1.4', 'sqlparse 0.3.0', 'tollalchamu' 1.2 1', 'scontrib-atchle 1.0.2', 'sphinxcontrib-websupport 1.1.0', 'sphinxcontrib-set 'xcontrib-atchle 1.0.2', 'sphinxcontrib-ismath 1.0.1', 'sphinxcontrib-hulle 1.0.2' 'sphinxcontrib-ismath 1.0.1', 'sphinxcontrib-thle 1.0.2' 'sphinxcontrib-ismath 1.0.1', 'sphinxcontrib-thle 1.0.2' 'sphinxcontrib-ismath 1.0.1', 'sphinxcontrib-thle 1.0.2' 'sphinxcontrib-ismath 1.0.1', 'sphinxcontrib-thle 1.0.2' 'sphinxcontrib-websupport 1.1.0', 'sphinxcontrib-thle 1.0.2' 'sphinxcontrib-ismath 1.0.1', 'sphinxcontrib-thle 1.0.2' 'sphinxcontrib-ismath 1.0.1', 'sphinxcontrib-thle 1.0.2' 'sphinxcontrib-websuport 1.1.0', 'sphinxcontrib-websuport 1.1.0', 'sphinxcontrib-websuport 1.1.0', 'sphinxcontrib-websuport 1.1.0', 'sphinxcontrib-websuport 1.1.0', 'sphinxcontrib-websuport 1.1.0', 'sphinxc t-rf 0.14.9', 'scikit-optimize 0.5.2', 'scikit-learn 0.23.1', 'scikit-image 0.14.2' el.yaml 0.15.89', 'ruamel-yaml 0.15.87', 'rrplugins 1.2.2', 'rpy2 2.9.1', 'rope 0.1 g 1.3.3', 'retry 0.9.2', 'researchpy 0.1.7', 'requests 2.24.0', 'requests-toolbelt 0', 'querystring-parser 1.2.3', 'QtPy 1.7.0', 'qtconsole 4.4.3', 'QtAwesome 0.57',

Libraries and codes available to be invoked





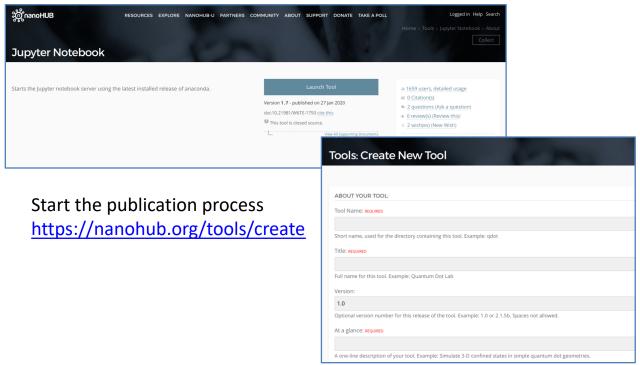
```
strachan@nanohub_1749163_77:~$ use 2>&1 | grep lammps
                 Large-scale Atomic/Molecular Massively Parallel Simulator
lammps-01Feb14:
lammps-03Mar20:
                 Large-scale Atomic/Molecular Massively Parallel Simulator
lammps-05Jun19:
                 Large-scale Atomic/Molecular Massively Parallel Simulator
                 Large-scale Atomic/Molecular Massively Parallel Simulator
lammps-06Apr15:
lammps-07Aug19Curl: Large-scale Atomic/Molecular Massively Parallel Simulator
lammps-07Aug19:
                 Large-scale Atomic/Molecular Massively Parallel Simulator
lammps-07Jul09:
                 Large-scale Atomic/Molecular Massively Parallel Simulator
lammps-09Dec14:
                 Large-scale Atomic/Molecular Massively Parallel Simulator
lammps-11Aug17:
                 Large-scale Atomic/Molecular Massively Parallel Simulator
lammps-12Feb07:
                 Large-scale Atomic/Molecular Massively Parallel Simulator
lammps-15Jan10:
                 Large-scale Atomic/Molecular Massively Parallel Simulator
lammps-15May15: * Large-scale Atomic/Molecular Massively Parallel Simulator
lammps-17Feb12:
                 Large-scale Atomic/Molecular Massively Parallel Simulator
lammps-22Aug18:
                 Large-scale Atomic/Molecular Massively Parallel Simulator
lammps-22Jan08:
                 Large-scale Atomic/Molecular Massively Parallel Simulator
lammps-300ct14:
                 Large-scale Atomic/Molecular Massively Parallel Simulator
lammps-31Mar17:
                 Large-scale Atomic/Molecular Massively Parallel Simulator
strachan@nanohub 1749163 77:~$ man use
strachan@nanohub 1749163 77:~$
```



Jump in!

Start developing your code!

https://nanohub.org/tools/jupyter (you will need a free nanoHUB account to launch Jupyter)



Share your work with your students and colleagues



Why deploy tools in nanoHUB?

- Tools are publications: indexed by Web of Science and Google Scholar.
- Impact statistics: usage metrics, secondary citations
- Anyone can run your tool without downloading/installing any software
 - Use it in the classroom
 - Document your research and make it reproducible (tools are containerized)
- A full scientific software development environment
 - Libraries for machine learning, statistics, data science, visualization, etc.
- Launch physics-based, research grade codes
 - LAMMPS, Quantum Espresso, NEMO5, abinit, MOOSE, and many more

