The MAterials Simulation Toolkit for Machine Learning (MAST-ML): Automating Development and Evaluation of Machine Learning Models for Materials Property Prediction

Ryan Jacobs, Tam Mayeshiba, Ben Afflerbach, Dane Morgan
(University of Wisconsin – Madison, WI USA)

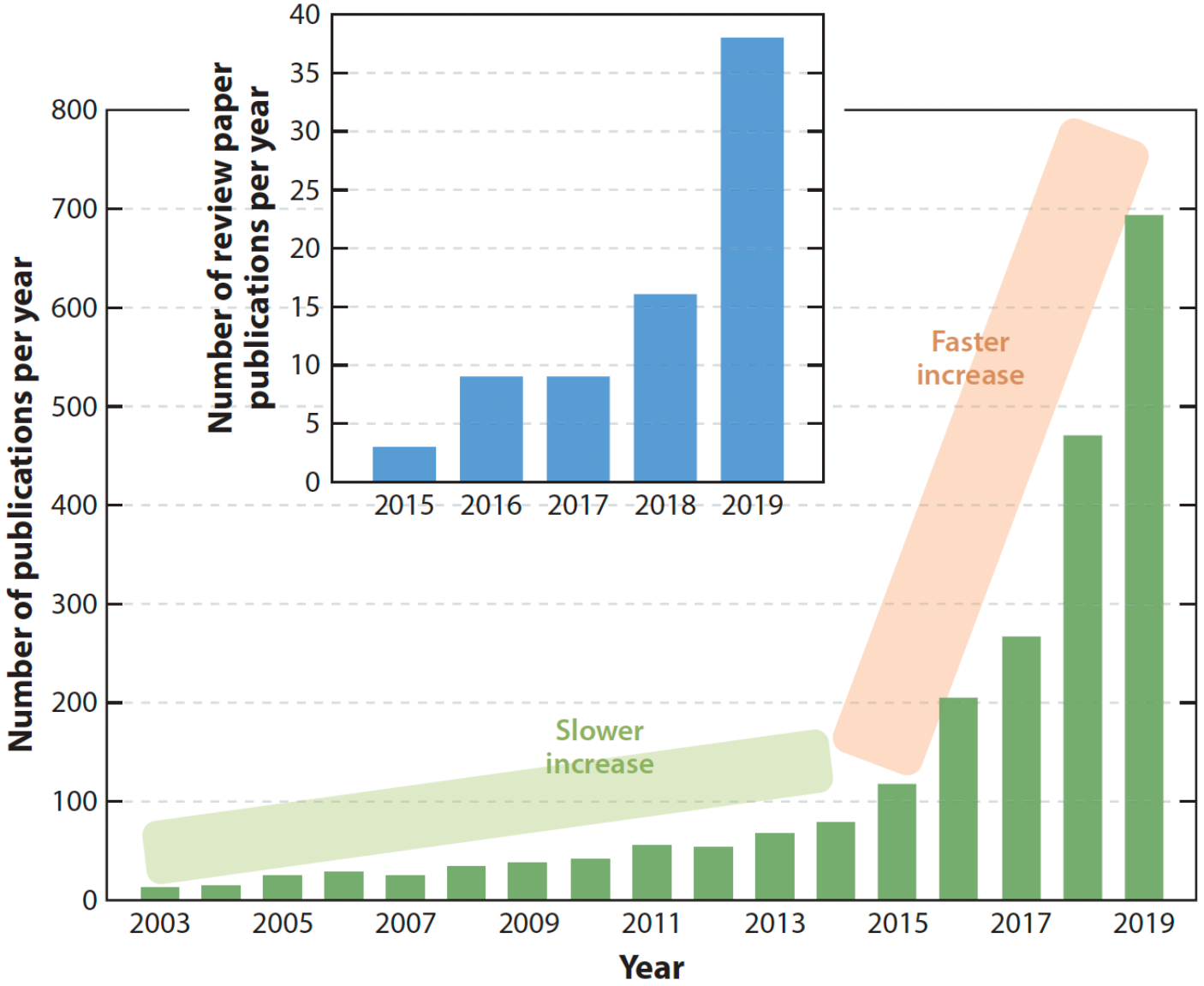
Luke Miles, Max Williams, Matthew Turner, Raphael Finkel
(University of Kentucky, Lexington, KY USA)

Most Recent Skunkworks MASTML members:
Avery Chan, Hock Lye Lee, Min Yi Lin

https://github.com/uw-cmg/MAST-ML

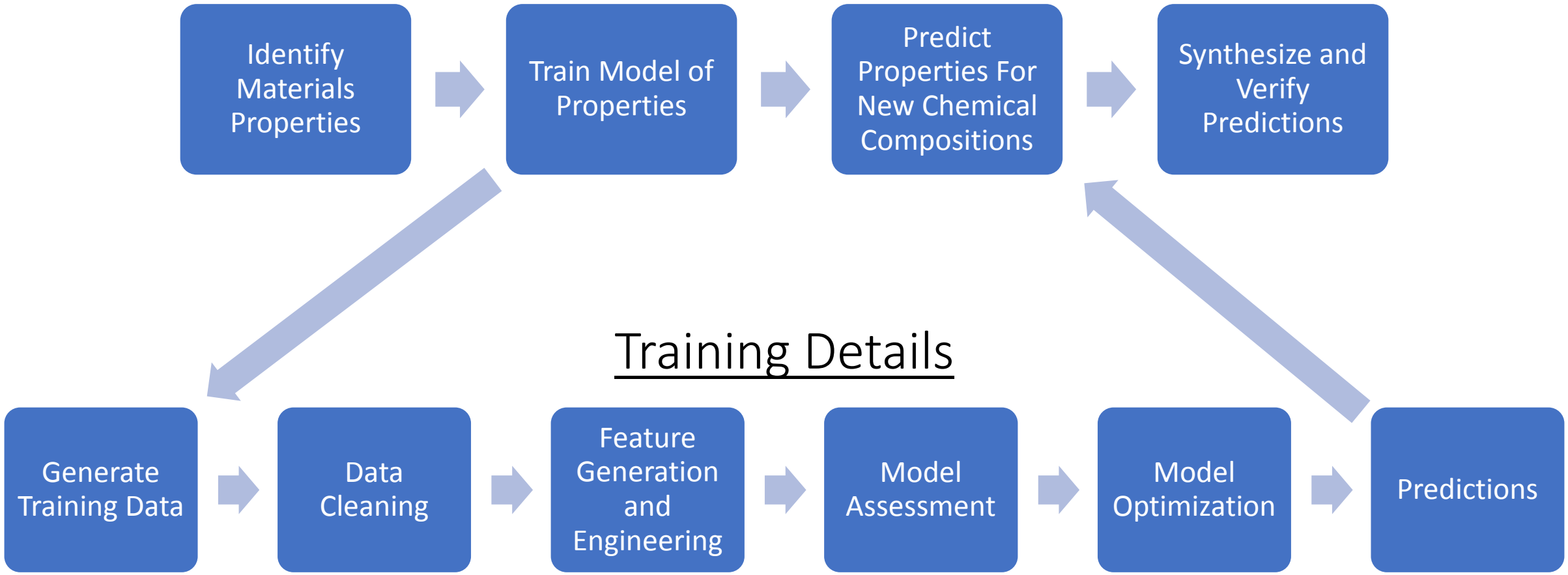
NanoHub ML Workshop
5/19/2021

Machine learning in Materials Science is Exploding



Jacobs and Morgan, Ann. Rev. Mat. Res. (2020), https://doi.org/10.1146/annurev-matsci-070218-010015

A Basic Materials Design Workflow



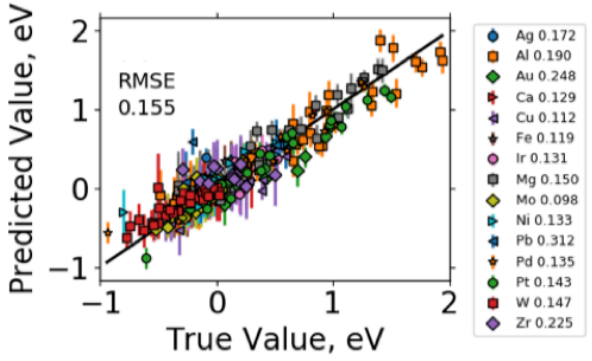
What is MAST-ML?

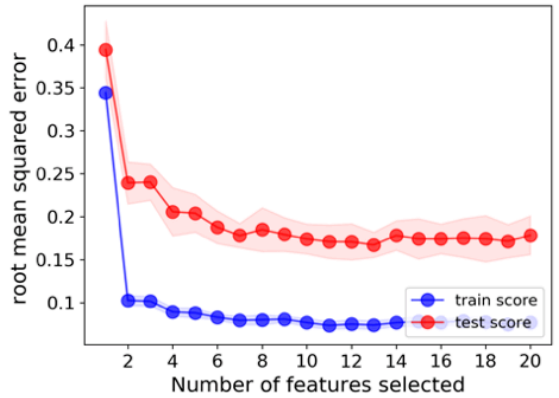
MAST-ML is an open-source Python package designed to broaden and accelerate the use of machine learning in materials science research, particularly for non-experts.

MAST-ML
MAterials Simulation Toolkit

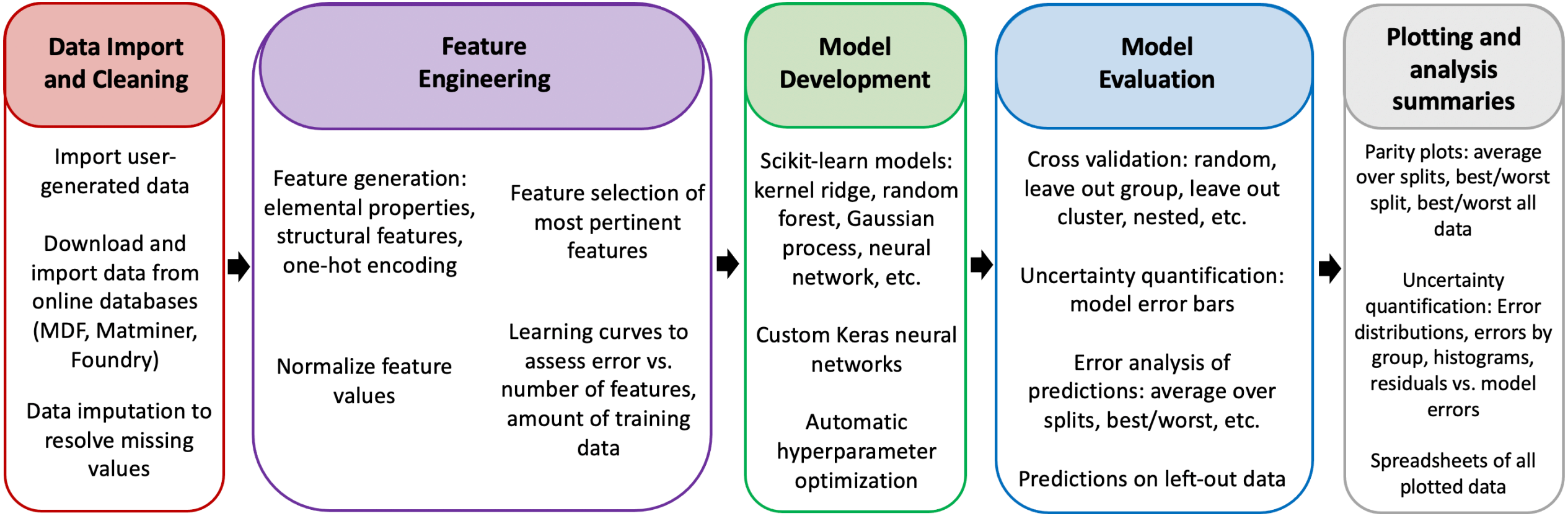
Automated machine learning tools for materials informatics research (MAST-ML)

https://github.com/uw-cmg/MAST-ML





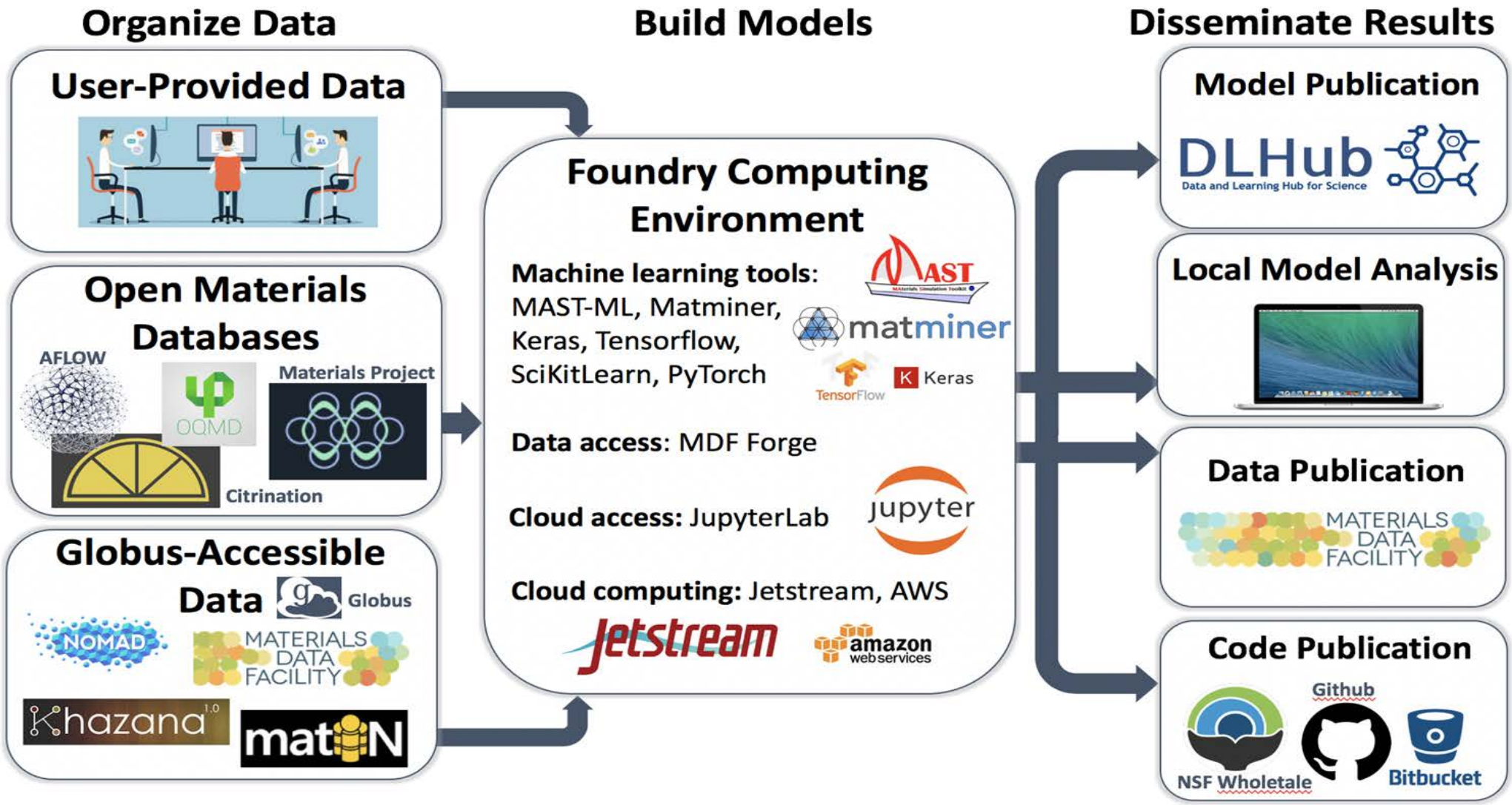
MAST-ML automates the supervised learning workflow



- MAST-ML supports the full library of scikit-learn modules, and can be used to construct neural networks with Keras (based on tensorflow)

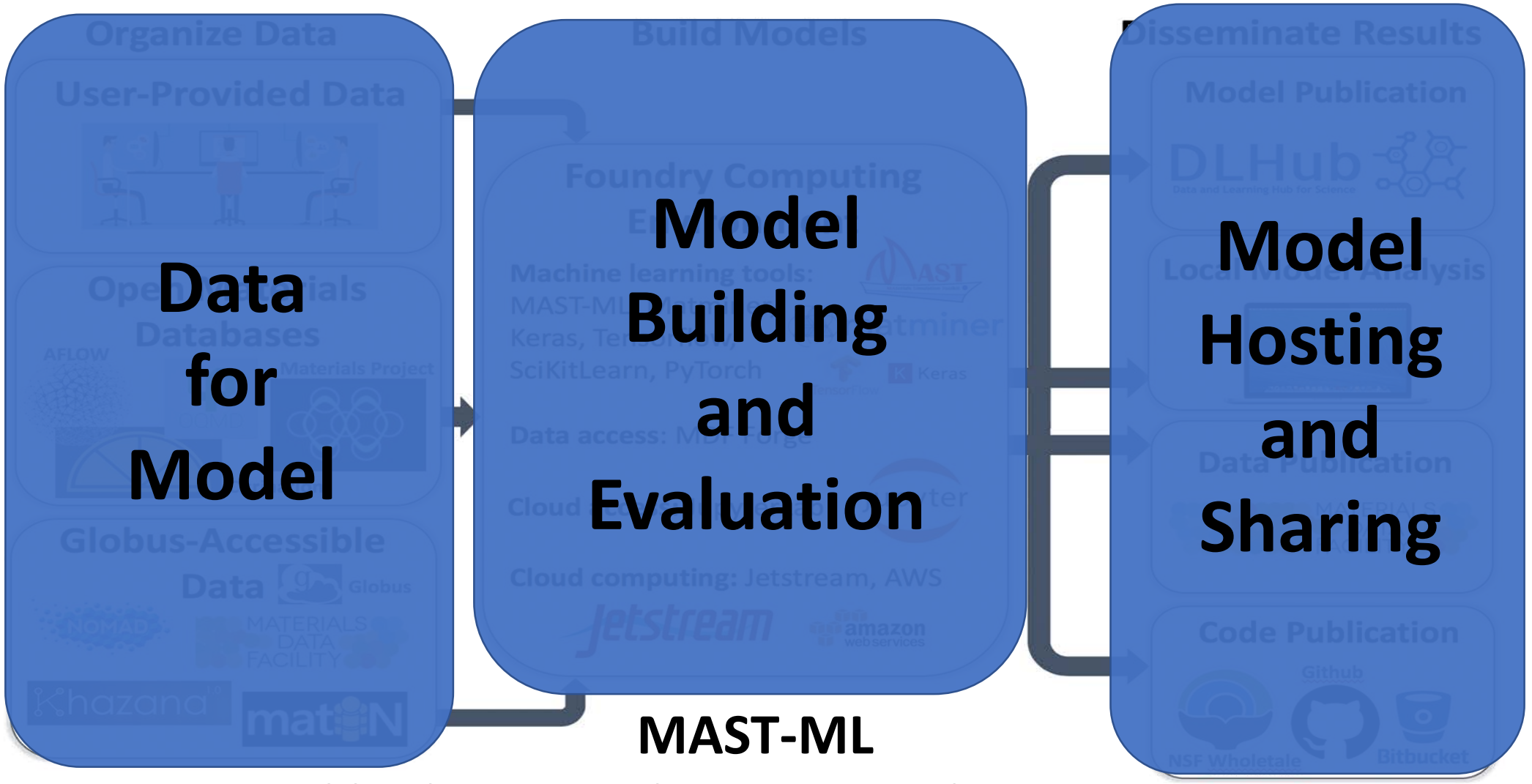
- MAST-ML allows for the simultaneous execution of an arbitrary combination of data preprocessing, feature generation/selection, model types and model evaluation metrics

(NSF CSSI) Machine Learning Materials Innovation Infrastructure



(PIs Dane Morgan, Paul Voyles, Michael Ferris, Ryan Jacobs, Ben Blaiszik)

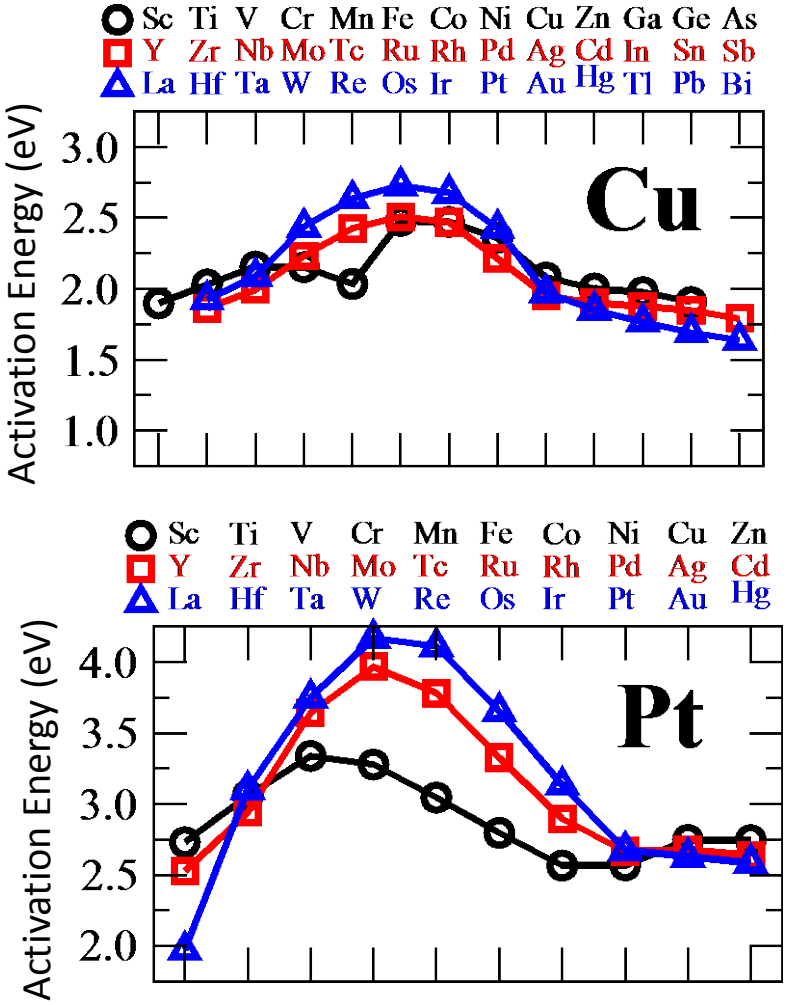
(NSF CSSI) Machine Learning Materials Innovation Infrastructure



Model building, evaluation, and key connections between data and model dissemination

Test Problem: Impurity Diffusion Database

- Diffusion of dilute impurity X in host H. We have DFT calculations of 440 values, but want ~4,000. [1, 2]
- Assume Y= Activation energies measured relative to host, X= Host descriptors, Impurity descriptors. Find Y=F(X).
- Descriptors = elemental properties like melting temperature, bulk modulus, electronegativity, … and their ratios, differences, etc. (MAGPIE set)[3]
- F is determined using standard machine learning regression methods (e.g., Gaussian Process Regression (Gaussian Kernel) (GPR), Random Forest (RF), neural network).
- Fit F with calculated data (15 hosts, 440 M-X pairs)



http://diffusiondata.materialshub.org/

[1] H. Wu, et al., *Comp. Mat. Sci* '17; [2] H. Lu, et al., *Comp Mat Sci* '19; [3] L. Ward, et al. *npj Comp. Mat.* '16

Getting Started with the MAST-ML tutorial on NanoHub

- Link to Tool: https://nanohub.org/tools/mastmltutorial
- Select “Launch Tool”
- A Jupyter notebook environment will open (may take a minute)
- Click on cell and run with Shift+return
- Data will be saved to local directory, see next slides for how to download results