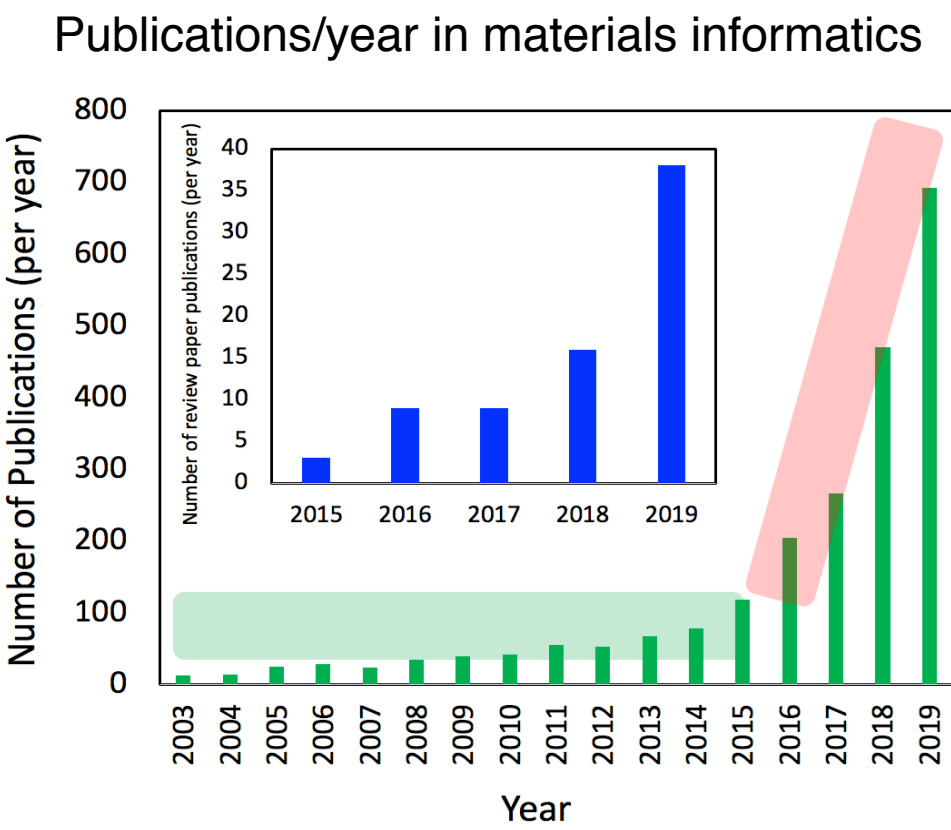# CSSI Framework: Machine Learning Materials Innovation Infrastructure

PI: Dane Morgan[1], Co-Pis: Ryan Jacobs[1], Paul Voyles[1], Michael Ferris[1], Ben Blaiszik[2]
Institutions: [1]University of Wisconsin, Madison, WI, [2]University of Chicago, IL

Machine Learning in Materials Science

Publications/year in materials informatics

- Machine Learning (ML) in materials science has exploded in recent years[1].
- Applications: Image processing, property database development, materials design, text mining of published papers, interatomic potentials, *ab initio* functionals, autonomous experiments

Goal: Support rapid development of machine learning applications in MS&E through (i) easy access to data, (ii) cloud-based tools for application of ML, and (iii) support for human and machine accessible and sustainable access to disseminated ML models.

Machine Learning Foundry Environment

Infrastructure for Model Use and Sharing

Data and Learning Hub for Science (DLHub)[2]

- Collect, publish, categorize models and associated code
- Operate models as a service to simplify sharing, consumption, and access
- Identify models with unique and persistent identifiers (e.g., DOI)
- Implement versioning, search, access controls etc.
- Publicly available:

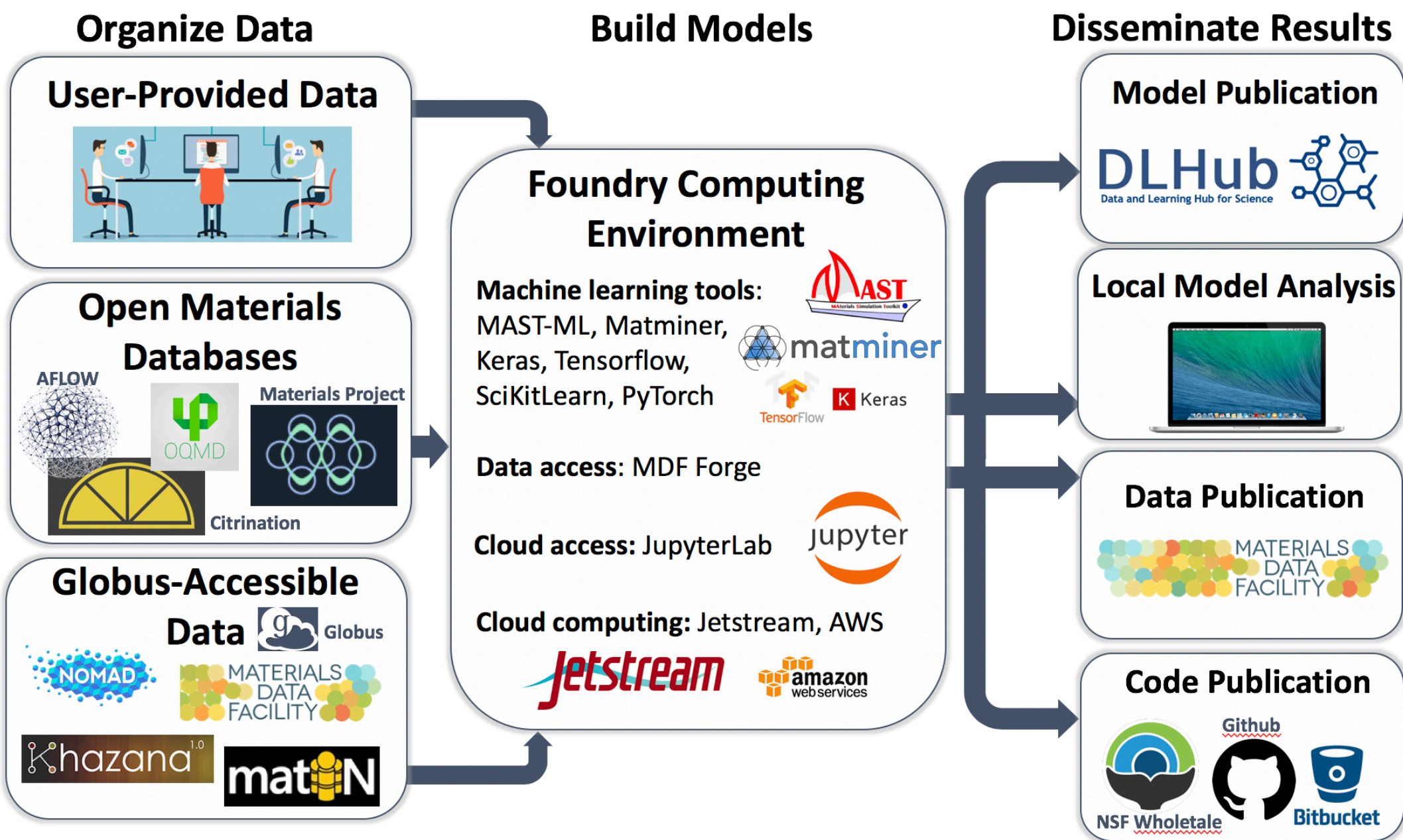https://github.com/DLHub-Argonne/dlhub_sdk
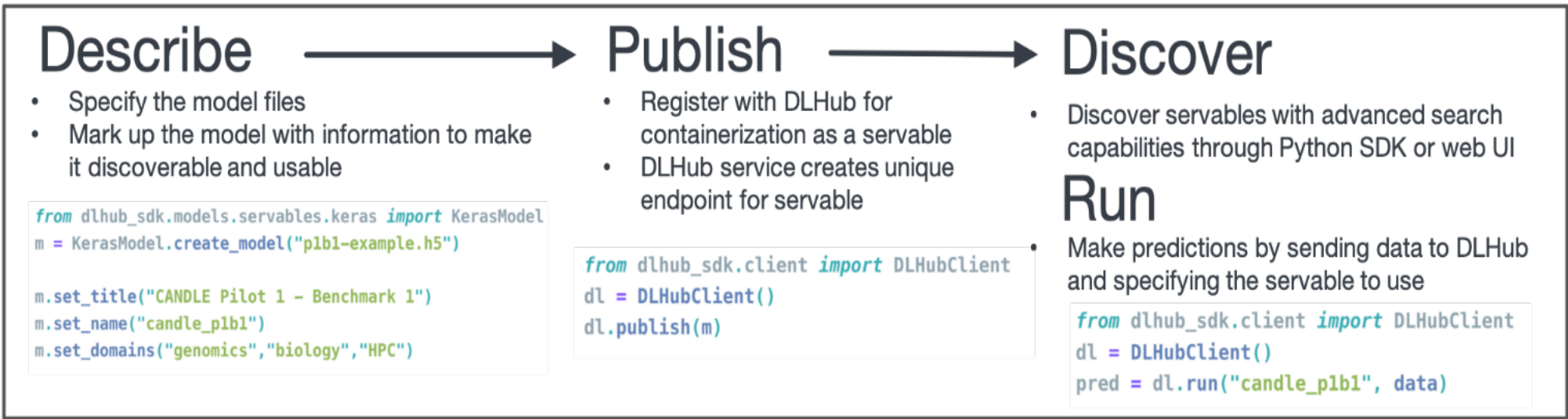
Impact Areas

Machine Learning Materials Innovation Infrastructure will:

- ***Transform*** the ability of materials researchers to apply machine learning
- ***Accelerate*** material discovery and design; enable new ML research modalities ***Support*** users to develop and disseminate ML models in a completely open-source environment
- ***Enable*** ML models to exist in a cloud-based ecosystem, where they can be used, shared, and updated easily through intuitive APIs
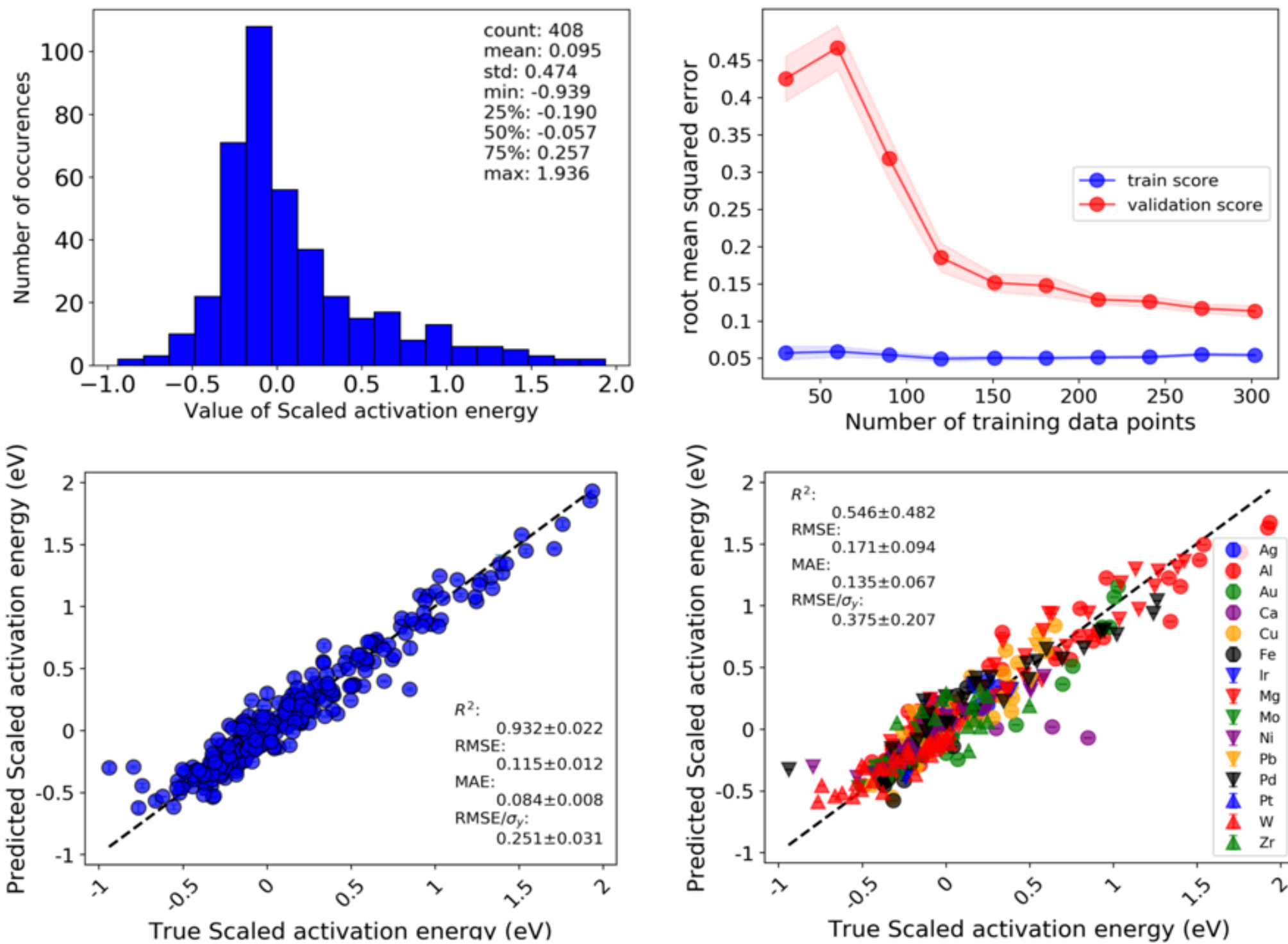
Automated Model Generation and Analysis

**Materials Simulation Toolkit for Machine Learning (MAST-ML)[3]**

- Automated tools for materials informatics
- Codifies best practices for development and analysis
- Easy to use, even for non-experts and non-programmers
- Publicly available: https://github.com/uw-cmg/MAST-ML

References

1. Morgan, D. and Jacobs, R. *Opportunities and Challenges for Machine Learning in Materials Science*. Annual Reviews of Materials Research (in review) (2020).
2. Chard, R., Li, Z., Chard, K., Ward, L., Babuji, Y., Woodard, A., Tuecke, S., Blaiszik, B., Franklin, M., Foster, I. *DLHub: Model and Data Serving for Science*. ArXiv: 1811.11213 (2018).
3. Jacobs, R., Mayeshiba, T., Afflerbach, B., Miles, L., Williams, M., Turner, M., Finkel, R., Morgan, D. *The Materials Simulation Toolkit for Machine Learning (MAST-ML): an automated open source toolkit to accelerate data-driven materials research*. ArXiv:1910.06291 (2019).

Broader Impact: The Informatics Skunkworks
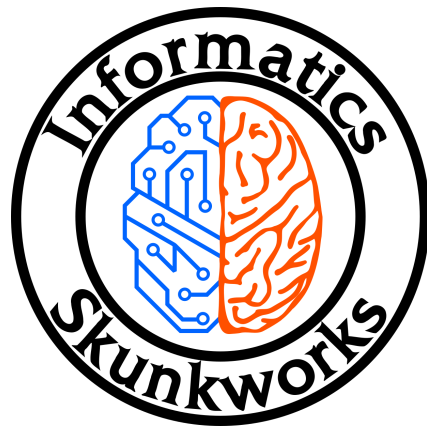
A program that engages undergraduates in authentic research at the intersection of informatics and science and engineering.

Over 150 students since 2015 have found

- Key skills: Research, Applied machine learning, Teamwork, Project management, Presentations
- Impactful science: 3 published papers, dozens of presentations
- Prestigious awards: >80% success rate
- Excellent Jobs/Graduate schools: MIT, Duke, Google, …

https://skunkworks.engr.wisc.edu/

Vision: A Community of Practice for Undergraduate Informatics Research