**MATDAT18: Materials and Data Science Hackathon**

**Time and Place**

May 15-17, 2018

NSF Headquarters, Alexandria, VA

**Organizers**

Brian Reich, Department of Statistics, North Carolina State University

Tim Mueller, Materials Science and Engineering, Johns Hopkins University

Sanguthevar Rajasekaran, Computer Science & Engineering, University of Connecticut

Andrew Ferguson, Materials Science and Engineering, University of Illinois

*Primary Contact*: [brian\_reich@ncsu.edu](mailto:brian_reich@ncsu.edu)

*MATDAT18 Webpage*: <https://matdat18.wordpress.ncsu.edu/>

**Synopsis**

Increases in computing power and advances in high-throughput instrumentation have led to the generation of computational and experimental materials science data sets of unprecedented size. Researchers are increasingly turning to data science tools to analyze these data to extract understanding and perform high-throughput screening and data-driven design. An impediment to success is that materials experts are infrequently also expert in data science, and data scientists typically lack the domain-specific expertise in materials engineering. It is the goal of this 3-day “hackathon” to partner materials and data scientists within interdisciplinary teams to spark collaborative research partnerships. Materials researchers will develop fluency in statistical and machine learning, and data scientists will be exposed to data-centric problems in materials engineering. **Full financial support is available to all participants.**

**Application Instructions**

**Step 1 – Solicitation of data-centric projects from materials researchers.**

***Deadline: January 19, 2018***

Materials scientists interested in proposing a project for the hackathon should complete the attached application form and submit via email to Brian Reich ([brian\_reich@ncsu.edu](mailto:brian_reich@ncsu.edu)). A (non-exhaustive) list of sample projects is provided below as examples of possible topics.

**Step 2 – Release and advertisement of materials projects to data scientists.**

***Deadline: March 15, 2018***

The organizers will sort the submitted projects, host them online, and advertise to the data science community through conferences, publications, and workshops. Data science applicants should complete the online application through the webpage at <https://matdat18.wordpress.ncsu.edu/>. The organizers will pair teams, perform remote introductions, and support preliminary preparation and goal setting in advances of the hackathon.

**Support and Sponsors**

Full financial support is available for participant travel, accommodation, and all meals.

Supported by the National Science Foundation under Grant No. DMR-1748198. Jointly supported by the Division of Materials Research (DMR), Division of Information and Intelligent Systems (IIS), and Division of Mathematical Sciences (DMS).

**Example Topic Areas**

*Materials Science*

• Data-driven design of experiment and simulation

• Inverse data-driven materials design

• Machine learning of quantitative structure property relationship (QSPR) models

• Identifying descriptors of materials performance

• Identifying patterns in experimental data (e.g. micrographs).

• Dimensionality reduction, exploration, and exploitation of high-dimensional data sets

• Materials discovery in large-scale databases

*Data Science*

• Bayesian data analysis

• Creation of databases

• Data integration

• Data reduction techniques

• Feature selection

• High Performance techniques

• Machine learning

• Out-of-core algorithms

• Spatial statistics

• Text mining

• Uncertainty quantification

**MATDAT18: Materials and Data Science Hackathon**

**MATERIALS SCIENCE TEAM APPLICATION FORM**

**Complete and return via email to** [brian\_reich@ncsu.edu](mailto:brian_reich@ncsu.edu) **by 15 January 2018**

**Team Composition** (2 people max.)

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Department** | **Institution** | **Email** |
|  |  |  |  |
|  |  |  |  |

**Project Title**

**Project Synopsis** (approx. 100 words)

**Identified Data-Science Collaborative Need** (approx. 100 words)

**Data Origin and Access** (*data must be available and sharable with data science teams* – please

address: data source/origin, access privileges, sharing privileges)

**Project Description** (approx. 1.5 pages, plus figures and references; please describe data size, form, dimensionality, uncertainties, number of examples, etc.)