

APS Scientific Computation Seminar Series

Speaker: Ben Blaiszik
Computation Institute
The University of Chicago

Title: Materials Data Facility - Streamlined and Automated Data Sharing,
Discovery, Access, and Analysis

Date: Monday, April 17, 2017

Time: 1:00 p.m.

Location: 401/A1100

Hosts: Nicholas Schwarz and Brian Toby

Abstract:

The Materials Data Facility (MDF), a NIST-funded Materials Genome Initiative project, operates two cloud-hosted production services, data publication and data discovery. These MDF services are built to promote open data sharing, self-service data publication and curation, and encourage data reuse, layered with powerful data discovery tools. The data publication service simplifies the process of copying data to a secure storage location, assigning data a citable persistent identifier, and recording custom (e.g., material, technique, or instrument specific) and automatically-extracted metadata in a registry while the data discovery service will provide advanced search capabilities (e.g., faceting, free text range querying, and full text search) against the registered data and metadata. The MDF services empower individual researchers, research projects, and institutions to publish research datasets, regardless of size, from distributed storage; and interact with and discover published and indexed data and metadata via REST APIs to facilitate automation, and analysis.

This talk will include live demonstrations of the services including a look at the new data discovery interface (Web UI and API) to find materials data indexed from over 15 databases and repositories and combine disparate materials datasets from distributed locations to train a state-of-the-art machine learning model on the JetStream National Science and Engineering Cloud.