Contribution ID: **20** Type: **not specified**

# Data ex Machina: Machine Learning with Jets in CMS Open Data

*Thursday 16 January 2020 09:00 (20 minutes)*

In this talk, I explore unsupervised and supervised machine learning techniques using CMS Open Data. I introduce a metric between jets based on the earth (or energy) mover's distance: the "work"required to rearrange one event into the other. Using this metric, I will probe the metric space of jets using unsupervised methods. Further, training supervised jet classifiers directly on data can potentially overcome the problematic reliance on simulated training data. I apply weakly supervised methods to train quark/gluon classifiers directly on the data and probe what the machine has learned. To enable machine learning research for jet physics using real LHC data, this dataset of over one million jets is made publicly available along with corresponding simulation.

**Authors:** METODIEV, Eric (Massachusetts Institute of Technology); KOMISKE, Patrick (Massachusetts Institute of Technology); MASTANDREA, Radha (Massachusetts Institute of Technology); NAIK, Preksha (Massachusetts Institute of Technology); THALER, Jesse (MIT)

**Presenter:** METODIEV, Eric (Massachusetts Institute of Technology)

**Session Classification:** Decorrelation and Semi/Unsupervised approaches