

Machine-learning from heterogenous data: dream or reality?

Thursday, 22 September 2022 11:00 (45 minutes)

Veracity (uncertainty of data quality) and variety (heterogeneity of form and meaning of data) are two of the 4V challenges of Big Data. Both are issues for the FAIRness of materials-science results, concerning in particular, the interoperability, i.e., the “I” in FAIR. I will address what may enable us to use heterogenous data for machine learning, e.g. data from different sources or exhibiting different quality. I will introduce metrics for measuring data quality and propose methods of unsupervised learning to explore large data spaces.

Presenter: Prof. DRAXL, Claudia (Humboldt Universität)

Session Classification: Infrastructure and Best Practices