SUBMIT AN ABSTRACT FOR THE FOLLOWING TMS2022 SYMPOSIUM:

MATERIALS DESIGN

ICME Case Studies: Successes and Challenges for Generation, Distribution, and Use of Public/Pre-Existing Materials Datasets

The emergence of digital data, public data repositories, and machine learning enables a new paradigm of materials research where high-quality datasets can be published and then reused and reanalyzed by other research teams, perhaps enabling entirely different applications than originally intended. The release of publicly available datasets has accelerated in recent years, encompassing varied datatypes such as densely sampled experimental data (e.g., synchrotron spectra and 3D serial section reconstructions), large quantities of image data (e.g., microstructure micrograph libraries), literature reviews containing sparsely populated and diversely measured material properties, and high-throughput large-scale simulation databases. The availability of these datasets provides the potential for faster and more cost-effective materials research by reducing unnecessary duplication of effort and effective division of labor. Despite these opportunities, this mode of research faces several challenges, including insufficient or incorrectly recorded metadata, lean or biased sampling of the materials space limiting (re-) analysis, and cultural norms limiting data sharing and accessibility.

This symposium solicits abstract submissions from researchers who are engaging in this research paradigm to share their experiences of the opportunities and challenges. Research involving dataset creation and publication and research involving reuse/reanalysis of external datasets are equally of interest.

Relevant topics include, but are not limited to:

- Case studies reviewing the successes and challenges of providing and/or using public datasets
- The provision of adequate metadata for reuse, or the use of datasets in the face of limited metadata
- Utilizing lean datasets for model building when further data acquisition is not possible
- Merging disparate datasets into a single cohesive dataset
- Model validation using externally obtained, high-dimensional digital datasets
- Examples of large dataset quality assessment, cleaning, and curation
- Uncertainty quantification of ICME predictions from lean data
- The public release of machine learning models trained on proprietary data such that the propriety data is protected

ORGANIZERS
Stephen J. DeWitt, Oak Ridge National Laboratory
Vikas Tomar, Purdue University
James Edward Saal, Citrine Informatics
James A. Warren, National Institute of Standards and Technology

SYMPOSIUM SPONSORS
TMS Integrated Computational Materials Engineering Committee