

Data Are Your Worst Employees: Challenges nanoHUB and HUBzero Face in Working with Data

Michael Zentner
Purdue University

nanoHUB.org has reached a point where it serves more than 1.4 million visitors annually, runs more than 800,000 simulations annually, and has a proven large impact in both research and education. Recently the nanoHUB community has asked for nanoHUB to support not only simulation but also working with data resources. Like employees, data are resources necessary for getting the job done. Data serve essential roles for analyzing impact retrospectively, taking immediate tactical action, or strategically planning the future. Unlike employees, we do not get to choose our data. Rather, they or the situation that creates them are placed upon us; and it is our job to learn how to manage these data to extract the value they contain. This talk will provide a general introduction to nanoHUB, and then will focus on several traits of data that are analogous to those of bad employees, with examples from applications of nanoHUB and other HUBzero sites in the domains of nano materials safety, medical informatics, and hub usage information. A data exploration environment called IDENT will be discussed as an initial effort from nanoHUB to work with ill-structured and ill-behaved data. IDENT also serves as a graphical approach to experimental and simulation space exploration that can be used to drive underlying computations and experiment.