Names Matter:
Eliminate the False Dichotomy Between Big Data and HPC

Daniel A. Reed
dan-reed@uiowa.edu
Vice President for Research and Economic Development
University Computational Science and Bioinformatics Chair
University of Iowa, USA
http://www.hpcdan.org

Done any low performance computing? Analyzed any small data? Of course, you did, when you checked your email and text messages on your smartphone a few minutes ago, but you did not think about it in those terms. In fact, you ran a trans-petascale computation and staged data across a global set of fiber optically connected data caches when you searched for a coffee shop yesterday. But, none of us call it that – it's just a quick web search in the ill-defined and amorphous cloud to which we are connected via a global wireless infrastructure.

Any expert in journalism, communications or psychology will immediately and unhesitatingly affirm that names have extraordinary power. They shape our thinking, define our discourse and reinforce our biases and perspectives. We use names and descriptors to categorize people, organizations and objects; we use names to define the positive and the negative; and we use names to market, reward and punish. High-performance computing (HPC) and big data/machine learning are not exempt from the power of names and a "they could learn from us if they'd only listen" dichotomy.

We are trapped in a sematic cul-de-sac when we talk about high-performance computing and computational science. They are code words for physics-inspired numerical solution of partial differential questions; all too frequently, everything else is potentially suspect and inferior. The denotation of the word exascale says nothing about computing at all, but the connotation of FLOPS and batch-oriented HPC platforms is there in all of our minds. In turn, deep learning and big data conjure images of computer science types tuning neural nets and recommender systems in cloud data centers for targeted product marketing.

It's time to stop categorizing and name calling, time to end cultural and disciplinary silos. There is no hierarchy of intellectual purity. There are just ideas and people who can learn from one other. Dare I say it – death to HPC; death to big data. It's time to stop the religious wars over HPC and big data cultures and technologies and focus on informatics-mediated discovery and innovation. The goal of an HPC center should not be self-perpetuation, either of infrastructure or organization. Nor should a cloud service be restricted to commercial domains. That means starting with intellectual outcomes and deriving approaches rather than defining approaches and searching for feasible intellectual outcomes.

There are many technical ways to envision the continuum and fusion of traditional HPC and computational science and big data and machine learning. These include integration of steam-based workflows and just-in-time schedulers, containerization and software stack packaging for application-specific configuration, fine-grained parallelization of learning packages, learning for algorithm and software adaptation and tuning, custom ASIC designs, software defined networks

and storage (SDN and SDS), and renewable and energy efficient hardware design and configuration. There are an equally large number of social and economic approaches, including denominating informatics costs in currency to illuminate investment priorities, spot pricing for priority access, economics-driven infrastructure selection and deployment, and institutional and funding agency policies for data management.