

Presentation title: Taming the Big Data Dragon

Abstract:

Nearly every major scientific revolution in history has been driven by one thing: data. Today, the availability of Big Data from a wide variety of sources is transforming health and biomedical research into an information science, where discovery is driven by our ability to effectively collect, manage, analyse, and interpret data. New technologies are providing abundance levels of thousands of proteins, population levels of thousands of microbial species, expression measures for tens of thousands of genes, information on patterns of genetic variation at millions of locations across the genome, and quantitative imaging data—all on the same biological sample. These omic data can be linked to vast quantities of clinical metadata, allowing us to search for complex patterns that correlate with meaningful health and medical endpoints. Environmental sampling and satellite data can be cross-referenced with health claims information and Internet searches to provide insights into the impact of atmospheric pollution on human health. Anonymized data from cell-phone records and text messages can be tied to health outcomes data, helping us explore disease transmission networks. Realizing the full potential of Big Data will require that we develop new analytical methods to address a number of fundamental issues and that we develop new ways of integrating, comparing, and synthesizing information to leverage the volume, variety, and velocity of Big Data. Using concrete examples from our work, I will present some examples that highlight the challenges and opportunities that present themselves in today's data rich environment.