Politics and Dirty Data

By Randy Bean | Contributor | August 12, 2014

A highly esteemed long-time colleague of mine tells the story of his coming out of MIT after completing a PhD program in computer science in the early 1990's. Dr. Luminary, as I'll refer to him, had worked in the MIT labs on parallel processing techniques for managing massively large amounts of data. This was Big Data in the laboratory stage 25 years ago. Dr. Luminary's first job was as a senior software engineer with Thinking Machines, the early pioneer in massively parallel processing systems. The company was so hot that Al Gore conducted a much publicized visit in 1987 – this may have been the genesis of his claim to having invented the Internet. At the very least, we should credit Mr. Gore for being onto the information super highway early on.

Dr. Luminary was very excited to use revolutionary new technology to help large Fortune 1000 companies extract key business insights from massive volumes of information. But it became clear that the large corporate world follows its own rules. The real obstacle to breakthrough innovation and success was due not so much to a lack of powerful new technology, but to "politics and dirty data." As he puts it, "We spent six months trying to reach agreement on a definition of what is a customer." As a former database marketer, my experience is that's more typical than unusual.

This apocryphal story highlights some well-documented challenges when it comes to making meaningful use of data and analytics. The first challenge is often organizational ("politics"). Data typically is a shared asset that cuts across the organization from production to consumption, with many touch points and derivations along the way. Many organizations have been actively engaged in tracing the lineage of their data and establishing data governance processes and standards so that there are some "rules of the road" that guide how organizations manage data. These rules include who "owns" the data; who has ultimate responsibility; and what is the process by which decisions governing data definition and usage are made. However well-intentioned all parties may be, reaching consensus on issues surrounding data is frequently a thorny process.

The second challenge is often referred to as data's "dirty little secret." This specifically pertains to the tremendous time and effort required to transform data ("dirty data") into a usable asset that has meaningful business value. Whole companies and an entire industry have been built to respond to this need. It is the ongoing lament of many a data analyst who complains that they spend "20% of our time

on data analysis, but 80% of our time on accessing and preparing the data." And, herein lies the big attraction and promise of Big Data for many a corporation – the ability to bypass the hundreds of hours of up-front data engineering to access the data much sooner and more easily, for purposes of analysis and putting this information to good use. This is the state of affairs today, as firms make the transition from traditional data environments to accelerated Big Data "lakes" and "hubs."

Lynda Applegate is the long-tenured Sarofim-Rock Professor of Business Administration at Harvard Business School, where for over 25 years, she has been teaching executives about innovation and entrepreneurship. Prof. Applegate has seen firsthand how critical and central data has become to the modern corporation, as well as to the innovative new startup. "The emergence of data and analytics in general management practices over the past decades has significantly influenced how businesses operate. Innovative firms have developed deep and rich data and analytical capabilities to distinguish themselves from their competitors," Prof. Applegate notes.

There is no question that access to insightful and timely data is enabling businesses, government agencies, medical researchers, and professional sports teams to spot opportunities and to act with greater agility. The biggest challenges are still human. To paraphrase the 911 Commission, "We had all of the data. We just didn't share it effectively and put the picture together".

Alas, Thinking Machines was ahead of its time, and filed for bankruptcy in 1994. Its assets were later acquired by Sun Microsystems. Al Gore ascended to the vice presidency of the United States in 1992, but fell short in his bid for the Presidency in 2000. He is now a venture capitalist. And Dr. Luminary still counsels organizations on how to navigate organizational politics and dirty data. Sometimes, progress comes gradually.

Randy Bean is CEO and managing partner of consultancy NewVantage Partners.