# Business Intelligence

**Richard T. Herschel**
*Saint Joseph's University, USA*

## INTRODUCTION

Big data is one of the most commonly written about topics in todays press. Routinely we are bombarded with reports about how much more data there is, how much more is now able to be captured, how many new sources it comes from, and how it is being used to in new and novel ways at the expense of our privacy. As Bernard Marr (2015) asserts it is a topic that is discussed in boardrooms, business publications, and the mainstream media, because big data provides new insights into everything. Big data encompasses traditional sources of structured transaction data that is now supplemented by mass quantities of unstructured data. This data is processed by new, inexpensive, and faster hardware that is then scrutinized by new and more advance analytics that provide organizations with more in-depth insight into their operational environment than ever before.

The role of business intelligence [BI] is to seek value from data. Today, BI combines text, video, voice, location data, social media, and any other new source of data with traditional data sets in order to learn about, interact with, and predict what is happening so that the organization can respond as fast as possible to whatever it perceives is the opportunity that the data reveals. BI deals with imperfect data that is oftentimes ambiguous, but which is available on a vast scale. As a result, Mayer-Shonberger and Cukier (2014) assert that the effect is that the extraction of value from big data is analogous to a treasure hunt. That is, organizations are scrutinizing big data to learn what is happening, without necessarily needing to understand why. They argue that in a big data world, correlations supersede causality, because the data is simply used to discover patterns and correlations in the data that offer novel and invaluable insights. The more data you have the better the insights. The underlying premise for BI then becomes this: the more data an organization can capture, the better the data-driven probability of understanding what is happening, and the faster you can respond to this insight. This means then that actions taken in BI are often based on an organizational/system confidence level in the analytic assessment of what the data suggests without any clear understanding of the root cause.

Big data would appear to many to be more about systems, and less about people. Certainly people are important because they are themselves a major source of big data fodder and it is oftentimes people's behavior that big data is trying to affect. Nevertheless, big data also is dependent on people because people must inevitably be responsible for how data is used, how it is managed, and for the consequences of the decisions made when using it.

This paper is intended then to remind us that big data is not simply something that data, systems, and analytics make happen and that we are somehow divorced from it and not responsible for unintended consequences. Instead, prudence would require that since we have unleashed big data, we have to somehow insure to the best of our ability that if we can't control big data we can at least use common sense in how we approach it and manage it.

## BACKGROUND

The amount of data in our world has exploded exponentially such that data, especially unstructured data, is now referred to as "big data". Where

DOI: 10.4018/978-1-5225-2255-3.ch082

measures of data were once gradually evolving from megabytes to terabytes, the sudden phenomena of big data accelerated these measures to volumes expressed in petabytes (1,024 terabytes) or exabytes (1,024 petabytes). The new influx of data is derived from billions to trillions of records of millions of people—all from different sources (e.g. Web, sales, customer contact center, social media, mobile data and so on). The data is typically loosely structured and often incomplete.

Big Data is the natural result of four major global trends:

1. Mobile computing
2. Social networking
3. Cloud computing
4. Moore's Law [processing power doubles every 2 years]

A lot of big data is derived from cell phone traffic and social networks, with much of it being stored in the cloud. The amount of this information grows almost exponentially as people routinely interact with companies and each other via wireless and wired networks as a matter of course.

The data being transmitted and captured from these transactions has become a key basis of competition, underpinning new waves of productivity growth, innovation, and consumer surplus, according to research by MGI and McKinsey's Business Technology Office (Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, & Hung Byers, 2011). From the standpoint of competitiveness and the potential capture of value, big data has had a substantial impact. Today, organizations leverage data-driven strategies to innovate, compete, and capture value from deep and up-to-real-time information. Hence, leaders in every sector are confronted with the need to grapple with the implications of big data in a quest to capture and harness its potential value.

The significance of big data to an organization falls into two categories: analytical use, and enabling new products. Big data analytics can reveal insights hidden previously by data too costly to process, such as peer influence among customers, revealed by analyzing shoppers' transactions, social and geographical data. Being able to process every item of data in reasonable time removes the troublesome need for sampling and promotes an investigative approach to data, in contrast to the somewhat static nature of running predetermined reports. Hence, big data expands BI's data-driven decision-making opportunities.

In their book, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Business", Minelli, Chambers, and Dhiraj (2013) present significant evidence supporting the value and justification for businesses to be 'big data-driven'. Minelli et al. define big data as data that goes beyond the traditional limits of data along three dimensions: volume, variety, and velocity. Volume is important because in the past, business use cases and predictive analyses were restricted because the data volume utilized was limited due to storage or computational processing constraints. However, with big data technology removing these constraints and allowing the use of unstructured data and more transaction data for larger data sets, Minelli et al., assert that organizations can now discover more subtle patterns that can lead to targeted actionable decisions, or allow them to factor in more observations or variables captured over a longer period of time into predictive models. Variety of data refers to the different types of data now available for analyses. Instead of simply using structured text or numbers, unstructured text, audio, images, geospatial information, and internet data are now captured and able to be analyzed. Velocity is about the speed at which data is created, accumulated, ingested, and processed. Minelli et al., state that since this data is now readily available, organizations pursue technologies to help them immediately process this information for real-time analytics-based decision-making.

IBM's Institute for Business Value (2013) added a fourth dimension to describe big data called *veracity*, or the uncertainty of data. They argue that the need to acknowledge and plan for uncertainty is a dimension of big data that has