Volume 22, No. 2

conwaymanagement News

547 Amherst Street, Suite 106, Nashua, NH 03063 USA Tel: 603-889-1130 www.conwaymgmt.com

Is Bigger Actually Better?

Well, when it comes to "big data," it very well might be. But, of course, it all depends on our expectations, how we define it, and how we use it. As you'll read in this newsletter, there are a number of misconceptions about big data, and certainly a few pitfalls to avoid. And then there are 3 keys to making the best use of it…

Mary Jane Conway-King

Big Data

by: Sheila Julien, Senior Associate

Everyone seems to be collecting more data in a variety of ways at increasing speed and varying levels of complexity. Let's begin by sharing what we mean by the term, 'Big Data.' Wikipedia defines 'Big Data' as "the collection of data sets so large and complex that it becomes difficult to process using on hand database management tools or traditional data processing applications". Another definition describes 'big data' as the condition when the volume, velocity, variety, variability and complexity of the data exceed an organization's storage and computing capacity.

As we look to the future, 'Big Data' is only going to get bigger but it may or may not produce a good return on the investment we make in it. If we are not careful, 'Big Data' can become a 'Big Waste' of time, money and opportunity.

How Has Data Changed, that it Deserves the Name "Big Data"?

The cost of storing data has plummeted. Because it costs less, we want more of it. Today large companies, are able to store mind-blowing quantities of data. Facebook users upload 250 million photos per day. Wal-Mart processes one million transactions an hour, and from these transactions they collect 2.5 petabytes of information. Every hour! Furthermore, access to cheap data collection and storage is no longer limited to the Wal-Mart's of the world. Cloud computing, such as that offered by Amazon Web Services, provides even small and mid-sized companies affordable ways to store huge amounts of data.

To put a petabyte (10^{15}) into perspective… A petabyte is one quadrillion bytes. One petabyte is equivalent to 20 million 4 drawer file cabinets filled with text, or 13.3 years of HDTV video.

In large part because of the decline in cost per byte, organizations are collecting **much richer data**. Instead of simply collecting weekly sales quantities or sales by product category or SKU, many organizations want to know *who* made *what* purchases, and *when*, sometimes *in what combination* with other products, and *how* they made the purchase. By studying who, what, when, where, and how with rigorous enough analytics, we may be able to discern 'why' they purchase and thereby learn how to get more people to purchase.

Weekly reports have moved to daily reports and sometimes to hourly reports (or even more frequent) in some cases. And it is not just sales that could benefit from the explosion of data and analytics. Operations and production have developed much richer data analysis of machine performance. A decade or more ago, companies would track machine downtime. Now they can track much more specific information about a whole range of variables that can assist in managing the equipment reliability — arriving at the optimal preventive maintenance schedule or preparation for machine failure.

conwaymanagement News
547 Amherst Street, Suite 106, Nashua, NH 03063 USA Tel: 603-889-1130 www.conwaymgmt.com

Another dimension to Big Data is the **increased variety of data sources** to collect and analyze: tweets, YouTube videos, Facebook interactions, email, website interactions, phone calls, recorded conversations, images, and spreadsheets in the cloud. The challenge is how to compile information from these data types and understand it well enough to make better decisions. Much of this new data comes in forms that are very different from those that data professionals are most familiar with. Hence it brings big new challenges in terms of how to analyze it and deliver insights quickly to get the value and benefit from it.

Examples of How We are Using Big Data Today?

Data mining helps companies recognize much faster when customer preferences and practices shift. Supermarkets use this kind of data mining to help recognize rapidly developing trends in food preferences that enable them to react quickly to be the market leader in this space. The additional speed and sensitivity to customer behavior that is newly advanced with Big Data also enables a business to better evaluate the effects of recent advertising campaigns. Data analytics can identify impact of promoting one item on other products. For example, if we promote Cheerios, will it impact sales of bananas or milk?

Macy's has used big data in their merchandizing pricing optimization system to reduce the time to optimize pricing of 73 million items for sale from 27 hours to just over an hour. They are able to use data from customer interactions in real time to respond quickly to their customer's experience.

UPS transports 16.3 million packages per day for 8.8 million customers. It responds to 39.5 million tracking requests per day!

By developing tools and methods to study individual consumers' behaviors, some retailers are able to understand aspects of the customer relationship that they previously couldn't get at. They can predict what else a shopper may be interested in. Amazon pioneered this approach with book recommendations, informing you that customers who bought this book also purchased this other book and customers who looked at this can opener also checked out these five alternatives. Netflix has also made extensive use of Big Data to predict what movies you might like based on what you have selected already.

"Most businesses have made slow progress in extracting value from big data. And some companies attempt to use traditional data management practices on big data, only to learn that the old rules no longer apply".

— Dan Briordy, "Big Data: Harnessing a Game Changing Asset," Economist Intelligence.

The newly available technology makes it much easier to automatically capture, measure, and transfer data about equipment function. Some organizations use extensive historical data about equipment reliability to predict with confidence required preventative maintenance. Using this analytical method they have been able to drastically reduce their preventive maintenance schedule, saving 40-50,000 maintenance hours, improving scheduling and planning and greatly reducing expediting necessitated by unanticipated failures. By studying the effects of failure, they can determine which components are safe to 'run to failure.'

While most organizations are gathering and using a great deal of quantitative data, many are having more difficulty making good use of qualitative data. Mining comments or other qualitative data is producing results for several organizations. One organization is working hard to improve safety through a safety observation system. Every observation that identifies someone making unsafe moves must include a comment in the record. This organization received 20,000 comments in one month. Mining the data for key words turned up a high frequency of comments noting a particular unsafe data location.

One tool used for mining qualitative data for frequent references is the 'word cloud.' One of these can be easily created to create a picture of the main themes of a survey or a comments field. Most frequent words

will loom larger than the rest. (At the conclusion of this report is a word cloud of this article, made by pasting the entire article into www.Wordle.net. As you can see, the word 'data' makes a frequent appearance.)

We also use Big Data to determine how we are doing. Why are we at 27% in this market with this product? Where can we improve? How should we rationalize our SKUs — where to cut to make room?

Many people might remember when IBM's Watson 'competed' on Jeopardy in February 2011. To answer questions, Watson examined 200 million pages of information to produce 'meaningful' voice enabled responses. Now, 3 years later, Watson is 24 times faster, with a 2,400 % improvement in performance and 90% smaller. In January 2014, IBM announced the creation of the IBM Watson Group. "Because Watson can process information akin to how people think, it represents a major shift in an organization's ability to quickly analyze, understand and respond to Big Data. Its ability to answer complex questions posed in natural language with speed, accuracy and confidence is transforming decision making across a variety of industries." IBM has partnered with a range of healthcare organizations to help transform how medicine is practiced, paid for and taught, with the help of Watson-powered solutions.

Hazards

For all the actual and potential value from Big Data, businesses also face hazards.

The data will be either structured or unstructured. If it is unstructured, it is impossible to analyze and requires time, skill, and effort to create structure. If it is structured, we need to understand the process used to structure it. What assumptions were made? What biases were introduced? There is no such thing as 'raw data' in the sense of data without any externally imposed meaning or influence. Even the act of designing a repository to capture the data will influence the data and hence the conclusions. If we are not careful to understand the implications of decisions about how we will capture and structure the data, we are vulnerable to acting on unrecognized biases. We need methods that will make sense of qualitative information without risking the human intervention and distortion often necessary today.

Furthermore, to make the data actionable requires the time and attention of people who can understand analytical methods and constraints as well as the business applications for the potential analysis. These people are scarce and if we not careful, our organizations can tie up a lot of resources analyzing huge quantities of data without arriving at improvements as valuable as these resources could achieve through other projects. Data analytics can be a 'time eater.' Big Data may divert valuable resources from working on better opportunities.

The risk of diverting resources from more valuable activities is especially high in organizations that have not arrived at clarity of purpose for the data. They have reams of data and set their resources to analyzing it, without first figuring out what data they need, how they will get it, and how they will use it.

Quite a lot of the data — both qualitative and quantitative — has a tremendous amount of 'noise' in it. While Twitter and Facebook can provide data about rapidly developing trends, but because of all the noise the signal may be detectable only in hindsight. A business can invest a lot of time and resources trying to be the first to discern important information only to discover that they have been chasing noise.

In retail, more data enables a company to tailor its products, services, and especially its advertising directly at individual customers to best match their needs and interests. Targeting in this way is much more efficient for the customers as well as for the retailer, but collecting this detailed customer information is also far more intrusive. While Amazon and Netflix pull it off by being very upfront about their recommendations, in many situations the intrusiveness can be quite off-putting. In *The Power of Habit*, Charles Duhigg tells the story of how Target, by carefully tracking and studying consumer habits, developed a way to infer with 85% accuracy

whether someone is pregnant — even before the person has told friends and family! The knowledge is very useful for Target and also provides the newly expectant individual with very relevant coupons — but at a cost of being too intrusive into the private lives of customers. Now Target intersperses largely irrelevant coupons among the baby coupons to create the *perception* that they are not being intrusive.

There are also difficulties in ensuring the independence and accuracy of qualitative data. The qualitative data includes judgmental information and the full meaning may vary by customer. Often customer qualitative information is filtered by the sales force before the information is catalogued and analyzed. Every time someone touches or transforms qualitative information to make it more readily analyzed, they introduce the possibility of inadvertently or purposefully twisting the message slightly to 'better suit' the expected message. The risk to data accuracy is very real when people are asked to measure and report quality errors or other data that is used to evaluate their own performance.

> ***"It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts."***
>
> ***— Sherlock Holmes***

Another question is whether we are capturing the right data. Many organizations are filled with tons of 'results' measures and very little by way of 'process' measures. Unless the organization understands and measures the processes as well, results measures provide little value because they do not lead to action that can improve the result. For example, customer complaints are a trailing indicator. The data is most useful if it is traced back to the process that led to the complaint — so that process measures can be put into place. Tracking internal system failures before they affect the client can provide monitoring of quality close to the source and increase delivered quality.

And sometimes, if data does not conform to previously held opinions some people simply will not believe it. Instead they will pore through reams of data to find some data that better supports 'reality' as they see it.

MIT researchers, McAffee and Bryjolfsson, report exactly the same phenomenon in the field: "Too often we saw executives who spiced up their reports with lots of data that supported decisions they had already made using the traditional 'Highest-Paid-Person's Opinion' approach. Only afterward were underlings dispatched to find the numbers that would justify the decision."

Nate Silver points out in the introduction to his book, The Signal and the Noise, Why So Many Predictions Fail but Some Don't, when access to data explodes, our biases tend to amplify the noise faster than we can learn to discern the signal. It is essential to spend our time on the 'signal' and figure out how to filter out the noise.

The Key to Better Use of Bigger Data

When thousands and thousands of answers are available through data analysis, the advantage goes to the organization with the better questions. If we do not start with the questions we need to answer, we will tie up a great deal of our resources with very little return on the investment. It matters little how cheap the data is to gather and store, if we do not transform it into actionable beneficial insights, decisions, and execution. And the costs of achieving this transformation are not declining nearly as fast as the costs of acquiring and storing the data.

Then, once you have identified the right questions, make sure you have accurate and sound data to answer them.

Second, ensure data integrity. Remember that the data does not tell us what the assumptions are. We need to dig in and understand the assumptions before we can conclude that we understand what the data tells us.

And third, get everyone on board with the idea that their mission is to understand and act on the data. Only when everyone truly is acting in alignment with the idea that the data can help us learn and improve, will the Big Data produce real results. When we let the data speak for itself and are ready to listen, learn, challenge, and relearn, we can reap the benefits of Big Data.