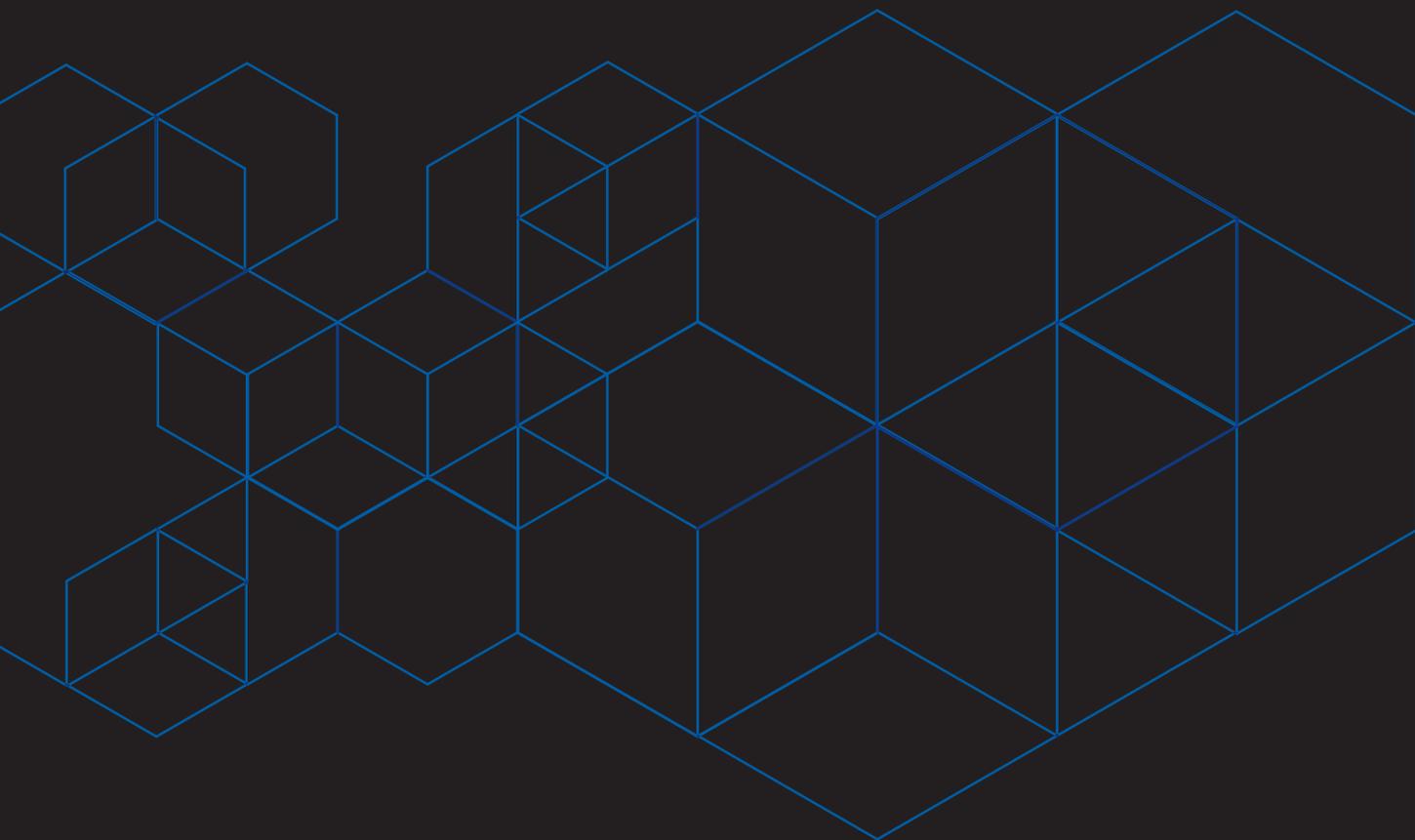


THE PROMISE AND CHALLENGE OF BIG DATA

A HARVARD BUSINESS REVIEW INSIGHT CENTER REPORT



Sponsored by



THE PROMISE AND CHALLENGE OF BIG DATA

The HBR Insight Center highlights emerging thinking around today's most important business ideas. In this Insight Center, we'll focus on what senior executives need to know about the big data revolution.

- 1 **Big Data's Management Revolution**
by Erik Brynjolfsson and Andrew McAfee
- 2 **Who's Really Using Big Data**
by Paul Barth and Randy Bean
- 3 **Data Is Useless Without the Skills to Analyze It**
by Jeanne Harris
- 4 **What Executives Don't Understand About Big Data**
by Michael Schrage
- 5 **Big Data's Human Component**
by Jim Stikeleather
- 6 **Will Big Data Kill All but the Biggest Retailers?**
by Gary Hawkins
- 8 **Predicting Customers' (Unedited) Behavior**
by Alex "Sandy" Pentland
- 9 **The Military's New Challenge: Knowing What They Know**
by Chris Young
- 10 **Three Questions to Ask Your Advanced Analytics Team**
by Niko Karvounis
- 11 **Metrics Are Easy; Insight Is Hard**
by Irfan Kamal
- 12 **Ignore Costly Market Data and Rely on Google Instead? An HBR Management Puzzle**
by Simeon Vosen and Torsten Schmidt
- 13 **Can You Live Without a Data Scientist?**
by Tom Davenport
- 14 **How to Repair Your Data**
by Thomas C. Redman
- 15 **What If Google Had a Hedge Fund?**
by Michael Schrage
- 16 **Get Started with Big Data: Tie Strategy to Performance**
by Dominic Barton and David Court
- 17 **Big Data's Biggest Obstacles**
by Alex "Sandy" Pentland
- 18 **To Succeed with Big Data, Start Small**
by Bill Franks
- 19 **The Apple Maps Debate and the Real Future of Mapping**
by Dennis Crowley
- 20 **Why Data Will Never Replace Thinking**
by Justin Fox
- 21 **Big Data Doesn't Work If You Ignore the Small Things That Matter**
by Robert Plant
- 22 **What Should You Tell Customers About How You're Using Data?**
by Niko Karvounis
- 23 **Data Can't Beat a Salesperson's Best Tool**
by Rick Reynolds
- 24 **When Pirates Meet Advanced Analytics**
by Robert Griffin
- 25 **What Could You Accomplish with 1,000 Computers?**
by Dana Rousmaniere
- 26 **Webinar Summary: What's the Big Deal About Big Data?**
featuring Andrew McAfee

BIG DATA'S MANAGEMENT REVOLUTION

BY ERIK BRYNJOLFSSON AND ANDREW MCAFEE

Big data has the potential to revolutionize management. Simply put, because of big data, managers can measure and hence know radically more about their businesses and directly translate that knowledge into improved decision making and performance. Of course, companies such as Google and Amazon are already doing this. After all, we expect companies that were born digital to accomplish things that business executives could only dream of a generation ago. But in fact the use of big data has the potential to transform traditional businesses as well.

We've seen big data used in supply chain management to understand why a carmaker's defect rates in the field suddenly increased, in customer service to continually scan and intervene in the health care practices of millions of people, in planning and forecasting to better anticipate online sales on the basis of a data set of product characteristics, and so on.

Here's how two companies, both far from being Silicon Valley upstarts, used new flows of information to radically improve performance.

Case #1: Using Big Data to Improve Predictions

Minutes matter in airports. So does accurate information about flight arrival times; if a plane lands before the ground staff is ready for it, the passengers and crew are effectively trapped, and if it shows up later than expected, the staff sits idle, driving up costs. So when a major U.S. airline learned from an internal study that about 10 percent of the flights into its major hub had at least a 10-minute gap between the estimated time of arrival and the actual arrival time — and 30 percent had a gap of at least five minutes — it decided to take action.

At the time the airline was relying on the aviation industry's longstanding practice of using the ETAs provided by pilots. The pilots made these estimates during their final approaches to the airport, when they had many other demands on their time and attention. In search of a better solution, the airline turned to PASSUR Aerospace, a provider of decision-support technologies for the aviation industry.

In 2001 PASSUR began offering its own arrival estimates as a service called RightETA. It calculated these times by combining publicly available data about weather, flight schedules, and other factors with proprietary data the company itself collected, including feeds from a network of passive radar stations it had installed near

airports to gather data about every plane in the local sky.

PASSUR started with just a few of these installations, but by 2012 it had more than 155. Every 4.6 seconds it collects a wide range of information about every plane that it "sees." This yields a huge and constant flood of digital data. What's more, the company keeps all the data it has gathered over time, so it has an immense body of multidimensional information spanning more than a decade. RightETA essentially works by asking itself, "What happened all the previous times a plane approached this airport under these conditions? When did it actually land?"

After switching to RightETA, the airline virtually eliminated gaps between estimated and actual arrival times. PASSUR believes that enabling an airline to know when its planes are going to land and plan accordingly is worth several million dollars a year at each airport. It's a simple formula: using big data leads to better predictions, and better predictions yield better decisions.

Case #2: Using Big Data to Drive Sales

A couple of years ago, Sears Holdings came to the conclusion that it needed to generate greater value from the huge amounts of customer, product, and promotion data it collected from its Sears, Craftsman, and Lands' End brands. Obviously, it would be valuable to combine and make use of all this data to tailor promotions and other offerings to customers and to personalize the offers to take advantage of local conditions.

Valuable but difficult: Sears required about eight weeks to generate personalized promotions, at which point many of them were no longer optimal for the company. It took so long mainly because the data required for these large-scale analyses was both voluminous and highly fragmented — housed in many databases and "data warehouses" maintained by the various brands.

In search of a faster, cheaper way, Sears Holdings turned to the technologies and practices of big data. As one of its first steps, it set up a Hadoop cluster. This is simply a group of inexpensive commodity servers with activities that are coordinated by an emerging software framework called Hadoop (named after a toy elephant in the household of Doug Cutting, one of its developers).

Sears started using the cluster to store incoming data from all its brands and to hold data from existing data warehouses. It then conducted analyses on the cluster directly, avoiding the time-consuming complexities of pulling data from various sources and combin-

ing it so that it can be analyzed. This change allowed the company to be much faster and more precise with its promotions.

According to the company's CTO, Phil Shelley, the time needed to generate a comprehensive set of promotions dropped from eight weeks to one and is still dropping. And these promotions are of higher quality, because they're more timely, more granular, and more personalized. Sears's Hadoop cluster stores and processes several petabytes of data at a fraction of the cost of a comparable standard data warehouse.

These aren't just a few flashy examples. We believe there is a more fundamental transformation of the economy happening. We've become convinced that almost no sphere of business activity will remain untouched by this movement.

Without question, many barriers to success remain. There are too few data scientists to go around. The technologies are new and in some cases exotic. It's too easy to mistake correlation for causation and to find misleading patterns in the data. The cultural challenges

are enormous and, of course, privacy concerns are only going to become more significant. But the underlying trends, both in the technology and in the business payoff, are unmistakable.

The evidence is clear: data-driven decisions tend to be better decisions. In sector after sector, companies that embrace this fact will pull away from their rivals. We can't say that all the winners will be harnessing big data to transform decision making. But the data tells us that's the surest bet.

This blog post was excerpted from the authors' upcoming article "Big Data: The Management Revolution," which will appear in the October issue of *Harvard Business Review*.

FEATURED COMMENT FROM HBR.ORG

Great synthesis of the biggest benefits to using big data. It's true; data has the potential to drive informed, real-time, and accurate communications. —GaryZ

11:00 AM SEPTEMBER 12, 2012

WHO'S REALLY USING BIG DATA

BY PAUL BARTH AND RANDY BEAN

We recently surveyed executives at Fortune 1000 companies and large government agencies about where they stand on big data: what initiatives they have planned, who's leading the charge, and how well equipped they are to exploit the opportunities big data presents. We're still digging through the data — but we did come away with three high-level takeaways.

- First, the people we surveyed have high hopes for what they can get out of advanced analytics.
- Second, it's early days for most of them. They don't yet have the capabilities they need to exploit big data.
- Third, there are disconnects in the survey results — hints that the people inside individual organizations aren't aligned on some key issues.

High expectations. Big data clearly has the attention of the C-suite — and responding executives were very optimistic for the most part. Eighty-five percent expected to gain substantial business and IT benefits from big data initiatives. When asked what they thought the major benefits would be, they named improvements in "fact-based decision making" and "customer experience" as #1 and #2. Many of the initiatives they had in mind were still in the early stages, so we weren't hearing about actual business results for the most part but rather about plans and expectations:

- 85 percent of organizations reported that they have big data initiatives planned or in progress.
- 70 percent report that these initiatives are enterprise-driven.

- 85 percent of the initiatives are sponsored by a C-level executive or the head of a line of business.
- 75 percent expect an impact across multiple lines of business.
- 80 percent believe that initiatives will cross multiple lines of business or functions.

Capabilities gap. In spite of the strong organizational interest in big data, respondents painted a less rosy picture of their current capabilities:

- Only 15 percent of respondents ranked their access to data today as adequate or world-class.
- Only 21 percent of respondents ranked their analytical capabilities as adequate or world-class.
- Only 17 percent of respondents ranked their ability to use data and analytics to transform their business as more than adequate or as world-class.

Notice that the bullet points above describe a set of increasingly sophisticated capabilities: gaining access to data, analyzing the various streams of data, and using what you've learned to transform the business. (Students of IT will recognize the familiar hierarchy: data must be transformed into information, and information must be transformed into knowledge.)

Problems with alignment? When we started to probe beneath the surface of these responses, we noticed that IT executives and line-of-business executives had quite different perceptions of their companies' capabilities. Some examples:

- How would you rate the access to relevant, accurate, and timely data in your company today? World-class or more than adequate — IT, 13 percent; business, 27 percent.
- How would you rate the analytical capabilities in your company today? World-class — IT, 13 percent; business, 0 percent.
- How would you rate your company on leaders' ability to use data and analytics to improve or transform the business? Less than adequate — IT, 57 percent; business, 18 percent.

To some extent these responses simply reflect a proximity bias: IT executives have a higher opinion of the company's analytical capability; similarly, business executives judge their own capacity to transform the business as higher than their IT colleagues do. But we suspect there's something else happening as well. Recall that 80 percent of respondents agreed that big data initiatives would reach across multiple lines of business. That reality bumps right up against the biggest data challenge respondents identified: "integrating a wider variety of data." This challenge appears to be more apparent to IT than to business executives. We'd guess that they're

more aware of how siloed their companies really are, and that this is another reason that they judge more harshly the company's capacity to transform itself using big data.

This disconnect continues when respondents rank the "current role of big data" in their company as planned or at proof of concept: only 31 percent of IT respondents felt the organization was at that stage, while 70 percent of the line-of-business executives thought they were at this stage.

Finally, in spite of the gap in perceptions, 77 percent of organizations report that there is a strong business/IT collaboration on big data thought leadership. This is probably too optimistic, from what we've seen when working inside companies and based on the gap in perceptions we saw in our survey. Job #1 is to get the organization aligned. Without that groundwork, big data can't live up to its promise.

FEATURED COMMENT FROM HBR.ORG

[This is an outstanding post. The issue around big data is tremendous. —Bruno Aziza](#)

9:00 AM SEPTEMBER 13, 2012

DATA IS USELESS WITHOUT THE SKILLS TO ANALYZE IT

BY JEANNE HARRIS

Do your employees have the skills to benefit from big data? As Tom Davenport and DJ Patil note in their October Harvard Business Review article on the rise of the data scientist, the advent of the big data era means that analyzing large, messy, unstructured data is going to increasingly form part of everyone's work. Managers and business analysts will often be called upon to conduct data-driven experiments, to interpret data, and to create innovative data-based products and services. To thrive in this world, many will require additional skills.

Companies grappling with big data recognize this need. In a new Avanade survey, more than 60 percent of respondents said their employees need to develop new skills to translate big data into insights and business value. Anders Reinhardt, head of global business intelligence for the VELUX Group — an international manufacturer of skylights, solar panels, and other roof products based in Denmark — is convinced that "the standard way of training, where we simply explain to business users how to access data and reports, is not enough anymore. Big data is much more demanding on the user." Executives in many industries are putting plans into place to beef up their workforces' skills. They tell me what employees need to become.

Ready and willing to experiment: Managers and business analysts

must be able to apply the principles of scientific experimentation to their business. They must know how to construct intelligent hypotheses. They also need to understand the principles of experimental testing and design, including population selection and sampling, in order to evaluate the validity of data analyses. As randomized testing and experimentation become more commonplace in financial services, retail, and pharmaceutical industries, a background in scientific experimental design will be particularly valued.

Google's recruiters know that experimentation and testing are integral parts of their culture and business processes. So job applicants are asked questions such as "How many golf balls would fit in a school bus?" or "How many sewer covers are there in Manhattan?" The point isn't to find the right answer but to test the applicant's skills in experimental design, logic, and quantitative analysis.

Adept at mathematical reasoning: How many of your managers today are really "numerate" — competent in the interpretation and use of numeric data? It's a skill that's going to become increasingly critical. VELUX's Reinhardt explains that "Business users don't need to be statisticians, but they need to understand the proper usage of statistical methods. We want our business users to understand how to interpret data, metrics, and the results of statistical models."

Some companies out of necessity make sure that their employees

are already highly adept at mathematical reasoning when they are hired. Capital One's hiring practices are geared toward hiring highly analytical and numerate employees in every aspect of the business. Prospective employees, including senior executives, go through a rigorous interview process, including tests of their mathematical reasoning, logic, and problem-solving abilities.

Able to see the big (data) picture: You might call this “data literacy”: competence in finding, manipulating, managing, and interpreting data, including not just numbers but also text and images. Data literacy skills must spread far beyond their usual home, the IT function, and become an integral aspect of every business function and activity.

Procter & Gamble's CEO, Bob McDonald, is convinced that “data modeling, simulation, and other digital tools are reshaping how we innovate.” And that has changed the skills needed by his employees. To meet this challenge, P&G created “a baseline digital skills inventory that's tailored to every level of advancement in the organization.” At VELUX, data literacy training for business users is a priority. Managers need to understand what data is available and to use data visualization techniques to process and interpret it. “Perhaps most important, we need to help them imagine how new types of data can lead to new insights,” notes Reinhardt.

Tomorrow's leaders need to ensure that their people have these

skills along with the culture, support, and accountability to go with them. In addition, they must be comfortable leading organizations in which many employees, not just a handful of IT professionals and PhDs in statistics, are up to their necks in the complexities of analyzing large, unstructured, and messy data.

Here's another challenge: the prospect of employees downloading and mashing up data brings up concerns about data security, reliability, and accuracy. But in my research, I've found that employees are already assuming more responsibility for the technology, data, and applications they use in their work. Employees must understand how to protect sensitive corporate data. And leaders will need to learn to “trust but verify” the analyses of their workforce.

Ensuring that big data creates big value calls for a reskilling effort that is at least as much about fostering a data-driven mind-set and analytical culture as it is about adopting new technology. Companies leading the revolution already have an experiment-focused, numerate, data-literate workforce. Are you ready to join them?

FEATURED COMMENT FROM HBR.ORG

This is a very interesting and timely post ... I am seeing the challenge of big data inducing increasing levels of anxiety right across all business sectors. —Nick Clarke

9:00 AM SEPTEMBER 14, 2012

WHAT EXECUTIVES DON'T UNDERSTAND ABOUT BIG DATA

BY MICHAEL SCHRAGE

How much more profitable would your business be if you had free access to 100 times more data about your customers? That's the question I posed to the attendees of a recent big data workshop in London, all of them senior executives. But not a single executive in this IT-savvy crowd would hazard a guess. One of the CEOs actually declared that the surge of new data might even lead to losses because his firm's management and business processes couldn't cost-effectively manage it.

Big data doesn't inherently lead to better results.

Although big data already is — and will continue to be — a relentless driver of revolutionary business change (just ask Jeff Bezos, Larry Page, or Reid Hoffman), too many organizations don't quite grasp that being big data-driven requires more qualified human judgment than cloud-enabled machine learning. Web 2.0 juggernauts such as Google, Amazon, and LinkedIn have the inborn advantage of being built around both big data architectures and cultures. Their future success is contingent upon becoming disproportionately more valuable as more people use them. Big data is both

an enabler and by-product of “network effects.” The algorithms that make these companies run need big data to survive and thrive. Ambitious algorithms love big data and vice versa.

Similarly, breakthrough big data systems such as IBM's Watson — the Ken Jennings-killing Jeopardy champion — are designed with a mission of clarity and specificity that makes their many, many terabytes intrinsically indispensable.

By contrast, the overwhelming majority of enterprise IT systems can't quite make up their digital minds. Is big data there to feed the algorithms or to inform the humans? Is big data being used to run a business process or to create situational awareness for top management? Is big data there to provide a more innovative signal or a comfortable redundancy? “All of the above” is exactly the wrong answer.

What works best is not a C-suite commitment to “bigger data,” ambitious algorithms, or sophisticated analytics. A commitment to a desired business outcome is the critical success factor. The reason my London executives evinced little enthusiasm for 100 times

more customer data was that they couldn't envision or align it with a desirable business outcome. Would offering 1,000 times or 10,000 times more data be more persuasive? Hardly. Neither the quantity nor quality of data was the issue. What matters is how — and why — vastly more data leads to vastly greater value creation. Designing and determining those links are the province of top management.

Instead of asking “How can we get far more value from far more data?” successful big data overseers seek to answer “What value matters most, and what marriage of data and algorithms gets us there?” The most effective big data implementations are engineered from the desired business outcomes in rather than from the humongous data sets out. Amazon's transformational recommendation engines reflect Bezos' focus on superior user experience rather than any innovative emphasis on repurposing customer data. That's real business leadership, not petabytes in search of profit.

Too many executives are too impressed — or too intimidated — by the bigness of the data to rethink or revisit how their organizations really add value. They fear that the size of the opportunity isn't worth the risk. In that regard, managing big data — and the ambitious algorithms that run it — is not unlike managing top talent. What compromises, accommodations, and judgment calls will you consider to make them all work well together?

Executives need to understand that big data is not about subordinating managerial decisions to automated algorithms but about deciding what kinds of data should enhance or transform user experiences. Big data should be neither servant nor master; properly managed, it becomes a new medium for shaping how people and their technologies interact.

That's why it's a tad disingenuous when Google-executive-turned-Yahoo-CEO thought leader Marissa Mayer declares that “data is apolitical” and that her old company succeeds because it

is so (big) data-driven. “It all comes down to data. Run a 1 percent test [on 1 percent of the audience], and whichever design does best against the user-happiness metrics over a two-week period is the one we launch. We have a very academic environment where we're looking at data all the time. We probably have somewhere between 50 and 100 experiments running on live traffic, everything from the default number of results for underlined links to how big an arrow should be. We're trying all those different things.”

Brilliant and admirable. But this purportedly “apolitical” perspective obscures a larger point. Google is a company with products and processes that are explicitly designed to be data-driven. The innovative insights flow not from the bigness of the data but from the clear alignment with measurable business outcomes. Data volume is designed to generate business value. (But some data is apparently more apolitical than others; the closure of Google Labs, for example, as well as its \$12.5 billion purchase of Motorola Mobility are likely not models of data-driven “best practice.”)

Most companies aren't Google, Amazon, or designed to take advantage of big data-enabled network effects. But virtually every organization that's moving some of its data, operations, or processes into the cloud can start asking itself whether the time is ripe to revisit their value creation fundamentals. In a new era of Watson, Windows, and Web 2.0 technologies, any organization that treats access to 100 times more customer data as more a burden than a breakthrough has something wrong with it. Big data should be an embarrassment of riches, not an embarrassment.

FEATURED COMMENT FROM HBR.ORG

[Big data is a means — not an end in itself. Being clear about the desired business outcomes is the start of employing big data to serve the business. —Pete DeLisi](#)

7:00 AM SEPTEMBER 17, 2012

BIG DATA'S HUMAN COMPONENT

BY JIM STIKELEATHER

Machines don't make the essential and important connections among data, and they don't create information. Humans do. Tools have the power to make work easier and solve problems. A tool is an enabler, facilitator, accelerator, and magnifier of human capability, not its replacement or surrogate — though artificial intelligence engines such as Watson and WolframAlpha (or more likely their descendants) might someday change that. That's what the software architect Grady Booch had in mind when he uttered that famous phrase “A fool with a tool is still a fool.”

We often forget about the human component in the excitement over data tools. Consider how we talk about big data. We forget that it is not about the data; it is about our customers having a deep,

engaging, insightful, meaningful conversation with us — if only we learn how to listen. So while money will be invested in software tools and hardware, let me suggest the human investment is more important. Here's how to put that insight into practice.

Understand that expertise is more important than the tool. Otherwise the tool will be used incorrectly and generate nonsense (logical, properly processed nonsense but nonsense nonetheless). This was the insight that made Michael Greenbaum and Edmund and Williams O'Connor — the fathers of modern financial derivatives — so successful. From the day their firm, O'Connor & Associates, opened its doors in 1977, derivatives were treated as if they were radioactive — you weren't allowed near them without a hazmat suit

and at least one PhD in mathematics. Any fool or mortgage banker can use a spreadsheet and calculate a Black-Scholes equation. But if you don't understand what is happening behind the numbers, both in the math and the real worlds, you risk collapsing the world financial system — or more likely your own business.

Understand how to present information. Humans are better at seeing the connections than any software is, though humans often need software to help. Think about what happens when you throw your dog a Frisbee®. As he chases it, he gauges its trajectory, adjusts for changes in speed and direction, and judges the precise moment to leap into the air to catch it, proving that he has solved a second-order, second-degree differential equation. Yeah, right.

The point is, we have eons of evolution generating a biological information processing capability that is different and in ways better than that of our digital servants. We're missing opportunities and risking mistakes if we do not understand and operationalize this ability.

Edward Tufte, the former Yale professor and leading thinker on information design and visual literacy, has been pushing this insight for years. He encourages the use of data-rich illustrations with all the available data presented. When examined closely, every data point has value, he says. And when seen overall, trends and patterns can be observed via the human "intuition" that comes from that biological information processing capability of our brain. We lose opportunities when we fail to take advantage of this human

capability. And we make mistakes. Tufte has famously attacked PowerPoint, which he argues overrides the brain's data-processing instincts and leads to oversimplification and inaccuracy in the presentation of information. Tufte's analysis appeared in the Columbia Accident Investigation Board's Report, blaming PowerPoint for missteps leading to the space shuttle disaster.

There are many other risks in failing to think about big data as part of a human-driven discovery and management process. When we over-automate big data tools, we get Target's faux pas of sending baby coupons to a teenager who hadn't yet told her parents she was pregnant or the Flash Crash on Thursday, May 6, 2010, in which the Dow Jones Industrial Average plunged about 1,000 points — or about 9 percent.

Although data does give rise to information and insight, they are not the same. Data's value to business relies on human intelligence, on how well managers and leaders formulate questions and interpret results. More data doesn't mean you will get "proportionately" more information. In fact, the more data you have, the less information you gain in proportion to the data (concepts of marginal utility, signal to noise, and diminishing returns). Understanding how to use the data we already have is what's going to matter most.

FEATURED COMMENT FROM HBR.ORG

Nice contextualization of the role that humans must play in the increasingly data-oriented world we are creating. —Jonathan Sidhu

7:00 AM SEPTEMBER 18, 2012

WILL BIG DATA KILL ALL BUT THE BIGGEST RETAILERS?

BY GARY HAWKINS

Increasingly, the largest retailers in markets across the country are employing sophisticated personalized marketing and thereby becoming the primary shopping destination for a growing number of consumers. Meanwhile, other retailers in those markets, once vigorous competitors for those loyalties, are being relegated to the role of convenience stores.

In this war for customers, the ammunition is data — and lots of it. It began with transaction data and shopper data, which remain central. Now, however, they are being augmented by demographic data, in-store video monitoring, mobile-based location data from inside and outside the store, real-time social media feeds, third-party data appends, weather, and more. Retail has entered the era of big data.

Virtually every retailer recognizes the advantages that come with better customer intelligence. A McKinsey study released in May 2011 stated that by using big data to the fullest, retailers stood to

increase their operating margins by up to 60 percent — this in an industry where net profit margins are often less than 2 percent. The biggest retailers are investing accordingly. dunnhumby, the analytics consultancy partnered with Kroger in the U.S. market, employs upwards of 120 data analysts focused on Kroger alone.

Not every retailer, however, has the resources to keep up with the sophisticated use of data. As large retailers convert secondary, lower-value shoppers into loyal, high-value shoppers, the growth in revenue is coming at the expense of competing retailers — all too often independent and mid-market retailers. This part of the retail sector, representing an estimated third of total supermarkets, has long provided rich diversity in communities across the United States. But it is fast becoming cannon fodder.

Within the industry, the term used for this new form of advantage is shopper marketing, loosely defined as using strategic insights into shopper behavior to influence individual customers on their paths

to purchase — and it is an advantage being bankrolled by consumer goods manufacturers’ marketing funds. A recently released study [PDF] by the Grocery Manufacturers Association (GMA) estimates annual industry spending on shopper marketing at more than \$50 billion and growing.

The growth in shopper marketing budgets comes as manufacturers are reducing the spending on traditional trade promotion that has historically powered independent retail marketing. Past retail battles were fought with mass promotions that caused widespread collateral damage, often at the expense of the retailer’s own margins. Today’s data sophistication enables surgical strikes aimed at specific shoppers and specific product purchases. A customer-intelligent retailer can mine its data searching for shoppers who have purchasing “gaps of opportunity,” such as the regular shopper who is not purchasing paper products, and targeting such customers with specific promotions to encourage them to add those items to their baskets next time they’re in the store.

A 2012 study by Kantar Retail shows manufacturer spending on trade promotion measured as a percentage of gross sales at the lowest level since 1999. But even this does not tell the whole story; it is the changing mix of manufacturer marketing expenditures that shows what is occurring. Trade promotion accounted for 44 percent of total marketing expenditures by manufacturers in 2011, lower than any other year in the past decade. This decrease is driven by a corresponding increase in shopper marketing expenditures.

As shopper marketing budgets have exploded, the perception has taken hold within the industry that a disproportionately large share of that funding is directed to the very largest retailers. That’s not surprising when you consider what Matthew Boyle of CNN Money reported recently. He noted that the partnership of Kroger and dunnhumby “is generating millions in revenue by selling Kroger’s shopper data to consumer goods giants . . . 60 clients in all, 40 percent of which are Fortune 500 firms.” It is widely understood that Kroger is realizing more than \$100 million annually in incremental revenue from these efforts.

The Kantar Retail report goes on to say “Manufacturers anticipate that changes in the next three years will revolve around continued trade integration with shopper marketing to maximize value in the face of continued margin demands. Manufacturers in particular expect to allocate trade funds more strategically in the future, as they shift to a ‘pay for performance’ approach and more closely

measure program and retailer performance.”

The same report calls out that the future success model will involve deeper and more extensive collaboration between the retailer and brand, with a focus on clear objectives and performance accountability. What needs to be recognized is that this manufacturer business model skews heavily to the capabilities of the largest retailers. It’s simply much easier for the brands to execute by deploying entire teams of people against a Safeway or a Target or a Walmart. It is much harder to interact with hundreds or thousands of independent retailers. Manufacturers’ past model of reaching independent retailers via wholesalers who aggregated smaller merchants for marketing purposes worked well in an age of mass promotion but not in an age of shopper-specific marketing. Wholesalers do not have shopper data and do not have sophisticated technologies or expertise in mining the data. Meanwhile, they have a challenging record of promotion compliance and in many cases lack the requisite scale for deep collaboration with brands.

Personalized marketing is proving to be a powerful tool, driving increased basket size, increased shopping visits, and increased retention over time. And if you’re one of the largest retailers, you get all these benefits paid for by CPG shopper marketing funds. But for everyone but those very large retailers, the present state of affairs is unsatisfactory. Independent retailers are keenly aware of the competitive threat and desperately want to engage, but they have had neither the tools nor scale to do so. The brand manufacturers are frustrated by increasing dependence on the very largest retailers even as they cave in to their inability to effectively and efficiently collaborate with a significant portion of the retail industry.

It would seem that the brand manufacturers’ traditional business model for marketing interaction with the independent retail sector is ripe for disruption. Growing consumer expectations of relevant marketing, the potential for gain if customer intelligence could be brought to the independent sector, and desire to mitigate the growing power of the largest retailers all provide powerful incentive to brand manufacturers. Independent retailers are savvy operators and are eager to join the fray if given the opportunity. Conversely, maintaining the status quo means the largest retailers continue to leverage personalized marketing to outpace smaller retailers, threatening the very diversity of the retail industry.

PREDICTING CUSTOMERS' (UNEDITED) BEHAVIOR

BY ALEX "SANDY" PENTLAND

Too often when we talk about big data, we talk about the inputs — the billions (trillions?) of breadcrumbs collected from Facebook posts, Google searches, GPS data from roving phones, inventory radio-frequency identification (RFIDS), and whatever else.

Those are merely means to an end. The end is this: big data provides objective information about people's behavior. Not their beliefs or morals. Not what they would like their behavior to be. Not what they tell the world their behavior is, but rather what it really is, unedited. Scientists can tell an enormous amount about you with this data. Enormously more, actually, than the best survey research, focus group, or doctor's interview — the highly subjective and incomplete tools we rely on today to understand behavior. With big data, current limitations on the interpretation of human behavior mostly go away. We can know whether you are the sort of person who will pay back loans. We can see if you're a good leader. We can tell you whether you're likely to get diabetes.

Scientists can do all this because big data is beginning to expose us to two facts. One, your behavior is largely determined by your social context. And two, behavior is much more predictable than you suspect. Together these facts mean that all I need to see is some of your behaviors and I can infer the rest just by comparing you to the people in your crowd.

Consequently, analysis of big data is increasingly about finding connections between people's behavior and outcomes. Ultimately, it will enable us to predict events. For instance, analysis in financial systems is helping us see the behaviors and connections that cause financial bubbles.

Until now, researchers have mostly been trying to understand things like financial bubbles using what is called complexity science or Web science. But these older ways of thinking about big data leave the humans out of the equation. What actually matters is how the people are connected by computers and how as a whole they create a financial market or a government, a company, or any

other social structure. They all can be made better with big data.

Because it is so important to understand these connections, Asu Ozdaglar and I have recently created the MIT Center for Connection Science and Engineering, which spans all the different MIT departments and schools. It's one of the very first MIT-wide centers, because people from all sorts of specialties are coming to understand that it is the connections between people that are actually the core problem in making logistics systems work well, in making management systems work efficiently, and in making financial systems stable. Markets are not just about rules or algorithms; they're about people and algorithms together.

Understanding these human-machine systems is what's going to make our future management systems stable and safe. That's the promise of big data, to really understand the systems that make our technological society. As you begin to understand them, then you can build better ones — financial systems that don't melt down, governments that don't get mired in inaction, health systems that actually improve health, and so much more.

Getting there won't be without its challenges. In my next blog post, I'll examine many of those obstacles. Still, it's important to first establish that big data is people plus algorithms, in that order. The barriers to better societal systems are not about the size or speed of data. They're not about most of the things that people are focusing on when they talk about big data. Instead, the challenge is to figure out how to analyze the connections in this deluge of data and come to a new way of building systems based on understanding these connections.

FEATURED COMMENT FROM HBR.ORG

I agree with the fact that big data is beginning to expose us to two facts: one, your behavior is largely determined by your social context, and two, behavior is much more predictable than you suspect.

— Anonymous

THE MILITARY'S NEW CHALLENGE: KNOWING WHAT THEY KNOW

BY CHRIS YOUNG

For soldiers in the field, immediate access to — and accurate interpretation of — real-time imagery and intelligence gathered by drones, satellites, or ground-based sensors can be a matter of life and death.

Capitalizing on big data is a high priority for the U.S. military. The rise in unmanned systems and the military's increasing reliance on intelligence, surveillance, and reconnaissance technologies have buried today's soldiers and defense professionals under a mountain of information. Since 9/11 alone, the amount of data captured by drones and other surveillance technology has increased a jaw-dropping 1,600 percent. And this avalanche of data will only increase, because the number of computing devices the armed services have in play is expected to double by 2020.

Rising to this challenge, defense companies have made major strides in image processing and analysis. Companies like our own have deployed technologies and software solutions for troops in Afghanistan that help soldiers quickly make sense of imagery and video feeds captured by unmanned systems flying overhead. And we are working on enhancing such technologies to decrease the lag time between gathering and interpreting data.

But even though advances are being made, the needs of military professionals are evolving as fast if not faster than the current pace of technology development can meet them. Keeping up will require defense companies to look beyond their own industry at the technology landscape as a whole.

To address soldiers' and diplomats' increasing need to understand both the cultural and geospatial context of their missions, for instance, defense companies need to become more adept at handling nontraditional sources of data such as social media. They need to find ways to quickly process this vast amount of information, isolate the most credible pieces of content, and quickly incorporate them with traditional intelligence sources such as video, overhead imagery, and maps. Defense contractors haven't had much experience tying rapid social media-processing tools into their existing systems, but they can draw lessons from other sectors in which significant technological advancements have been made. A great case in point is social analytics start-up BackType's real-time streaming and analytics tool.

The defense industry would also do well to learn from the rapid development processes that have made the technology sector so

operationally agile. Gone are the days when the Department of Defense was willing and able to routinely purchase high-risk concepts that exist only in PowerPoint presentations. With the slowdown in federal defense spending, government customers are looking for solutions that are mature and ready to be used in the field.

What's more, with government budgets under pressure, defense companies developing big data applications cannot count on sizeable government incentives. That means they will need to assume greater risk than in the past, not only in seeking to fulfill the military's current needs but also in strategically investing in the future. For companies like our own, with already-established data collection and processing businesses, the market opportunity makes the investment worth it and critical to long-term success.

Defense providers that are able to meet this challenge will not only be successful with their traditional defense customers but they will also find opportunities beyond the Pentagon. The rapid data-processing and analysis tools defense companies are developing to enable soldiers to quickly receive drone-captured intelligence could, for instance, be applied to the health care and emergency response fields. This technology could allow health professionals across different regions to pick up on trends and more quickly respond to medical epidemics such as West Nile virus and swine flu. Real-time image processing could also be tailored to help disaster response teams save more lives and better identify damage during hurricanes and other episodes of severe weather. The payoff cannot be understated.

The growing confluence of big data and national defense comes during a period of industry uncertainty and a shift in U.S. defense strategy and thinking. But just as the military is evolving to meet the demands of the twenty-first century, the defense industry must also adapt. This means being more nimble, more focused on anticipating customers' needs, and more attuned to developments in other sectors confronting big data. In the future, the government will be equipping soldiers with better and faster tools to prevail on a networked battlefield and increasingly across a hostile cyber landscape. These same applications also have the potential to change the way we interact with data on a daily basis. The defense industry has the opportunity and responsibility — not only to its customers but also to shareholders and employees — to take the lead and address this challenge.

THREE QUESTIONS TO ASK YOUR ADVANCED ANALYTICS TEAM

BY NIKO KARVOUNIS

Here's something that senior managers should keep in mind as they launch big data initiatives: *advanced analytics is mostly about finding relationships between different sets of data.* The leader's first job is to make sure the organization has the tools to do that.

Three simple, high-level questions can help you guide progress on that front — and keep people focused on that central task. In a later post, I'll propose a second set of questions that arise when the organization is deeper into its big data initiatives.

1. How are we going to coordinate multichannel data?

Businesses operate in more spheres than ever — in-store, in-person, telephonic, Web, mobile, and social channels. Collecting data from each of these channels is important, but so is coordinating that data. Say you're a manager at a consumer retail store — how many Web customers also purchase at your brick-and-mortar stores, and how often?

One solution here is a common cross-channel identifier. At Quovo we've built an investment analysis platform that aggregates investors' accounts from across multiple custodians and brokerages into one customer profile. This allows investors to easily run analyses on the full picture of their investments — no matter where the data is housed.

Ultimately, that's the value of a common identifier for any business: a fuller picture of related data under a single listing. In the retail example, a single registration account for Web and mobile commerce can help consolidate data from both channels in order to give a better picture of a customer's online shopping. Even more broadly, a customer loyalty program can help, because it gives consumers a unique ID that they apply to every purchase, regardless of the channel. Drugstores such as CVS and Walgreens have been using this system for years to track customer behavior and to get a full picture of purchasing patterns, loyalty trends, and lifetime customer value.

A final note: common identifiers are useful for any organization but may be particularly important for large organizations that manage multiple systems or have grown through acquisitions. In this case, shared identifiers can help bridge different data sets and systems that otherwise might have trouble "speaking" to each other.

2. How are we going to deal with unstructured data?

If your organization wants to get serious about fully mining the value of data, then addressing unstructured data is a must. Unstructured data is messy, qualitative data (think e-mails, notes, PDF statements, transcripts, legal documents, multimedia, etc.) that doesn't fit nicely into standardized quantitative formats. It

can hold rich insights — a commonly cited example being doctors' handwritten clinical notes, which often contain the most important information about patient conditions.

There are a few different ways to begin thinking about capturing unstructured data. Your database systems can have room for form fields, comments, or attachments; these allow unstructured sources and files to be appended to records. Metadata and taxonomies are also useful. Metadata is data about data — tagging specific listings or records with descriptions to help categorize otherwise idiosyncratic content. Taxonomies are about organizing data hierarchically through common characteristics. In the example of medical records, you could tag patient records showing high levels of cholesterol (this tag would be an example of metadata) and then set up your data governance to be able to drill down into this group by gender and within gender by age; the ability to support this increasing granularity within a category is an example of taxonomies.

3. How can we create the data we need from the data we have?

Ultimately, data analytics are useful only if they help you make smarter business decisions — but the data you have may not be as relevant to those decisions as it needs to be. Businesses need to think hard about which variables or combination of variables are the most salient for key business decisions.

Auto insurance providers deal with this issue every day, as I discovered during my work in the sector with LexisNexis. Today many insurance carriers are piloting telematics programs, which track policyholders' driving patterns in real time through in-car devices. This telematics data is then entered into actuarial models to predict driving risk (and thus insurance premiums). The idea is that direct driving behavior over time will be more predictive than traditional proxies such as age, credit rating, or geography. While this seems like a logical assumption, the real question isn't whether driving behavior is more predictive than traditional proxies but whether driving behavior combined with traditional proxies are most predictive of all.

For insurers, transforming this data into its most usable form may require the creation of new composite variables or scores from the existing data — something like a driving risk score that gives weight to telematics data, geography, and credit score. The beauty of this approach is that it consolidates multiple, unique data streams into one usable metric that speaks directly to a critical business decision — whom to insure and for how much. What's the equivalent of a driving score for your organization?

Big data is complicated stuff, and the three questions discussed here aren't the end of the road. But they do speak to the strategic

mind-set that senior managers must keep in order to get the most out of advanced analytics — and generate a rich, layered data con-

text around the messy realities of business.

12:00 PM SEPTEMBER 24, 2012

METRICS ARE EASY; INSIGHT IS HARD

BY IRFAN KAMAL

Big data is great. But we should consider that we've actually had more data than we can reasonably use for a while now. Just on the marketing front, it isn't uncommon to see reports overflowing with data and benchmarks drawn from millions of underlying data points covering existing channels such as display, e-mail, Web sites, searches, and shopper/loyalty — and new data streams such as social and mobile engagement, reviews, comments, ratings, location check-ins, and more.

In contrast to this abundant data, insights are relatively rare. Insights here are defined as actionable, data-driven findings that create business value. They are entirely different beasts from raw data. Delivering them requires different people, technology, and skills — specifically including deep domain knowledge. And they're hard to build.

Even with great data and tools, insights can be exceptionally tough to come by. Consider that improving Netflix's recommendation engine accuracy by about 10 percent proved so challenging that only two teams — of tens of thousands from more than 180 countries competing for the \$1 million prize — were able to hit the goal. Or that despite significant work to improve online display ad targeting, the average click-through rate (and, by implication, relevance) still remains so low that display ads on average receive only one click for every 1,000 views. That is, the vast majority of people who see the ad don't think it's interesting or relevant enough to click on.

When they are generated, though, insights derived from the smart use of data are hugely powerful. Brands and companies that are able to develop big insights — from any level of data — will be winners.

Here's a four-step marketing data-centered process that doesn't stop at the data but focuses instead on generating insights relevant to specific segments or affinity groups:

- 1. Collect.** Good data is the foundation for the process. Data can be collected from sources as varied as blogs, searches, social network engagement, forums, reviews, ad engagement, and Web site clickstream.
- 2. Connect.** Some data will simply be useful in the aggregate (for example, to look at broad trends). Other data, however, is more actionable if it's connected to specific segments or even individuals. Importantly, the linking of social/digital data to individuals will require obtaining consumer consent and complying with local regulations.

3. Manage. Given the speed and volume of social interaction online, simply managing big data requires special techniques, algorithms, and storage solutions. And while some data can be stored, other types of data are accessed in real time or for only a limited time via APIs.

4. Analyze and Discover. This part of the process works best when it's a broadly collaborative one. Using statistics, reporting, and visualization tools, marketers, product managers, and data scientists work together to come up with the key insights that will generate value broadly for specific segments of customers and ultimately personalized insights for individual customers.

Consider these insights — drawn from detailed studies and data analysis — that are being used by us and others to deliver value today:

Friends' interests make ads more relevant. Based on the evaluation of social graph data and clicks, companies such as 33Across have found that showing ads based on friends' similar interests can substantially raise ad click/conversion rates.

Sometimes it's okay if people hate your TV show. A television network commissioned Ogilvy to look at the relationship between social media buzz and ratings. An analysis of thousands of social media data points and Nielsen ratings across 80 network and cable shows identified ways to help predict ratings changes and find the specific plot lines and characters that could be emphasized in marketing to drive higher viewership. One insight was that it's critically important to look at data differently by show and genre. As an example, for some reality and newly launched cable shows, both love and hate — as long as there was lots of it — drove audience ratings.

Social media works best in combination. Measuring the actual business impact of social media and cross-media interactions (beyond just impressions) is in the early stages and could have perhaps the most profound impact of all on making marketing better and more efficient. For example, by exploring panel-based data on brand encounters by socially engaged customers in the restaurant industry, Ogilvy and ChatThreads found that social media was very effective in driving revenue in this segment. However, this effect was strongest when social media were combined with other channels such as traditional PR and out-of-home media. Exposure to these combinations drove increases of 1.5 to 2 times in the likelihood of revenue gains.

Each of these insights works because it is actionable and generates value. Each one provides a concrete road map for making marketing more effective and efficient. And applying each insight creates value that both brands and consumers can appreciate.

FEATURED COMMENT FROM HBR.ORG

[These are] good suggestions for achieving insights that actually create business value. —Guy Horst

9:00 AM SEPTEMBER 25, 2012

IGNORE COSTLY MARKET DATA AND RELY ON GOOGLE INSTEAD?

An HBR Management Puzzle

BY SIMEON VOSEN AND TORSTEN SCHMIDT

“It’s decision time,” Stefanie said to her younger colleague. “I can’t wait any longer. The managing director needs to know when and where we’re launching. Google Trends is telling us one thing. Our traditional market research is telling us something else. I need to decide which one to believe. And why are you smiling?”

“I’m sorry,” Eugen said, trying to keep a straight face. “It’s just that you’ve been having this argument with yourself for two months: should we or shouldn’t we trust Google’s free data on the intensity of Web searches? Look!” he suddenly exclaimed. “There’s one of those Rohloff motors. What a work of art — even though it’s one of our competitors. Fourteen gears spaced evenly at 13.6 percent. A 526 percent total gear range. Beautiful.”

Stefanie soon saw what her tech-obsessed colleague had noticed in the crowd of bicycles here along the Danube: a black electric motor attached to the frame of a handsome white bicycle. The rider of the e-bike drew admiring stares as he swiftly overtook other cyclists. Stefanie and Eugen had gone for a walk to finalize the details of their company’s launch of a similar high-end electric bicycle named the Oryx. These were exciting times for their employer, the German automaker Vado, as it moved into the rapidly expanding e-bike industry, but Stefanie, the manager of the new e-bike unit, was feeling the weight of the decisions she faced.

“I want to believe the Google data. I really do,” she said to Eugen, the Oryx’s chief engineer. “But it’s — it’s — Google.”

(Editor’s note: This fictional Management Puzzle dramatizes a dilemma faced by leaders in real companies. Like HBR’s traditional case studies, HBR.org’s Management Puzzles are based on academic research into business problems. This story was inspired by “Forecasting private consumption: survey-based indicators vs. Google trends” by Simeon Vosen and Torsten Schmidt in the Journal of Forecasting (September 2011). Please contribute by offering insights, solutions, and stories from your own experience.)

“The only thing quote unquote wrong with the Google data is that it’s free,” said Eugen. “That’s why I always smile when I hear you agonize over whether to rely on it. I think the reason you trust traditional market research is that you have to pay for it out of your

budget.” Stefanie was indebted to Eugen for his brilliance, but that grin was beginning to grate on her.

“The data may be free, but we still have to spend money on analysis,” she said.

“Still — it’s nothing compared to the cost of corporate market research.”

Stefanie was a new-product pro — she had handled numerous product launches when she was in Vado’s motorcycle unit. And of course she knew how to Google things on the Internet — who doesn’t? But Google Trends was a whole new world for her. She had stared wide-eyed as Eugen showed her spikes in consumer interest in e-bikes during specific times of the year — spring, when cycling enthusiasts were beginning to take to the roads, and midsummer, when the Tour de France was on every television and mobile phone — and in certain regions. Hungary, for instance. Massive amounts of data were there at her fingertips and all for free. She had spent some money on an outside firm that specialized in Google-data analysis, but the cost hadn’t amounted to much.

“When you pay real money for something, there’s accountability,” she said as she and Eugen crossed the long footbridge on their way back to the company’s executive building with its soaring profile visible in the distance. “Remember when our market research people showed us that in Germany, France, and Switzerland purchase intentions for e-bikes reach a peak in the weeks before Christmas? We were able to go back to them and ask them questions. How big was your sample? How do you know your data is accurate? For every question, we got an answer. Our market research people know what they’re doing. That’s why I used their data to set our prices, and it’s why I have confidence in their analysis about a launch date and site. They say November in Switzerland, and I’m inclined to agree with them. The Berne Bicycle Fair, roads with giant hills, and consumers who are fanatical about cycling and have money to spend. It all adds up.”

“But it turned out their sample was small and heavily skewed toward older consumers. There was a lot of spin in their results.”

“I admit the sample wasn’t huge,” she said. “With everyone using

mobile phones now instead of landlines, telephone surveys are difficult. But older consumers are one of our most important markets.”

“Google draws on millions and millions of searches, from people of all ages. Google sees that the real heat in the e-bike market comes in the early spring and is strongest in Eastern Europe. Google shows we should launch in March — in Budapest.”

“And if Google is wrong, what recourse do we have?” she asked.

“Google isn’t going to be wrong. Google data is like this river; it’s an organic part of nature. It’s not concocted like the analysis of a survey or a focus group. It comes directly from real users. It’s pure and unbiased, and it’s updated every week, so we can see trends as they develop. That’s why I keep arguing that Google Trends should be our go-to source of market data — not only just for this launch but also for next year’s launches in southern Europe and the United Kingdom.

And for all our decisions on where we put our advertising money?”

“But we don’t really know the exact relationship between the intensity of Google searches and product purchases, even with all the analysis we’ve done. And what if Google stops publishing this kind of data someday? Or starts charging for it?”

“That’s ‘someday.’”

They walked in silence. Finally Stefanie said, “If the managing director hears we’re basing our launch decision on Google data, he’ll be livid. He’ll want to know what’s the point of spending millions on traditional market research if we can get all the information we need for free off the Internet.”

That grin again. “Exactly,” Eugen replied.

Question: Should the company place its trust in the market data that Google provides for free?

11:00 AM SEPTEMBER 26, 2012

CAN YOU LIVE WITHOUT A DATA SCIENTIST?

BY TOM DAVENPORT

There are some great articles about big data and analytics in the current issue of HBR, and I am happy to have coauthored one of them with DJ Patil, one of the world’s first practicing data scientists. In fact, data scientists are what our article is about. I would argue that they are the most important resource for capitalizing on big data. While there is a lot of Hadoopalooza in the technology press about the tools for managing big data, and they are wonderful, it’s also true that they are (a) widely available and (b) mostly free. Neither can be said of data scientists. The other necessary resource, massive quantities of data, can also be found on every virtual corner these days. If your customers have Internet access, for example, you’ve got big data.

Simply put, you can’t do much with big data without data scientists. They are the magicians who transform an inchoate mass of bits into a fit subject for analysis. God may have been the first to produce order out of chaos, but data scientists do it too, admittedly on a smaller scale. They can suck data out of a server log, a telecom billing file, or the alternator on a locomotive and figure out what the heck is going on with it. They create new products and services for customers. They can also interface with carbon-based life forms — senior executives, product managers, CTOs, and CIOs. You need them.

But here’s the burning question: do you need them now and on your payroll? Data scientists are definitely in short supply at the moment; ask any company that’s trying to hire them. And trained data scientists seem to be aware of their scarcity, judging from the compensation packages they are pulling down. One admitted to me

in an interview that “We’re a pain in the ass. We’re constantly telling managers that their data isn’t good enough or interesting enough. We’re never satisfied with how the company is being run. We ask for big options packages, and then we leave after six months.”

I suspect that over the next couple of years it will become a lot easier to hire data scientists. A wide variety of universities — including Harvard, Berkeley, Stanford, Columbia, NC State, and others — are offering data science courses and starting degree programs. (Berkeley is reputed to be starting three different programs, which I’m not sure represents progress.) In the meantime, there are a number of consulting organizations that can offer data science services. The big analytics providers, including Accenture, Deloitte, and IBM, are all gearing up on the topic, and there are a variety of boutique firms that have formed. The large offshore providers, of which Mu Sigma is the largest, are also getting in on the big data game.

If you want to stake out a leadership position in big data for your company — and I would encourage you to do so — you don’t really have an alternative to joining the data scientist recruiting scrum. But if you want to dip a toe into the water, you might think about starting with a consultant. In either case, in the meantime start sending people to school on the topic.

FEATURED COMMENT FROM HBR.ORG

I agree ... [Taking shortcuts now will result in paying a high price later in the game.](#) —Maneesh Joshi

HOW TO REPAIR YOUR DATA

BY THOMAS C. REDMAN

One can't help being impressed with the effort biologists, physicists, and other scientists devote to data quality. From the careful design of experiments and data collection processes to explicit definitions of terms to comprehensive efforts to ensure that the data is correct, no effort is spared. This is not surprising. After all, data is the lifeblood of science.

Increasingly, data is also the lifeblood of business and government. And the attention science pays to data quality provides important lessons, especially for those interested in big data.

Simply put, bad data make everything about big data — from discovering something truly novel to building a product or service around that discovery to monetizing the discovery — more difficult. The two most important problems are:

1. The data is poorly defined, leading to incorrect interpretations.
2. The data is simply wrong, incomplete, or out of date, leading to problems throughout.

Worse, in business bad data can be downright dangerous. Consider that throughout the mid-2000s financial companies did a terrific job slicing, dicing, and packaging risk in creating collateralized debt obligations (CDOs). But they either didn't know or didn't care that too much of the mortgage data used to create them was wrong. Eventually, of course, the bad data asserted itself. And the financial system nearly collapsed.

Early computer programmers recognized that having bad data was an issue and coined the expression “garbage in, garbage out.” The big data update is “big garbage in, big TOXIC garbage out.”

This example and observation underscore a very critical point: no matter what, neither underestimate the data quality problem nor the effort required to solve it. You must get in front of data quality. You can systematically improve data by following these recommendations, inspired by the best scientific traditions and the efforts of leading companies to translate those traditions into business practice. To start, think of data quality problems as falling into two categories, each requiring a different approach.

Address preexisting issues. There are some problems that have been created already, and you have no choice but to address these before you use the data in any serious way. This is time-consuming, expensive, and demanding work. You must make sure you understand the provenance of all data, what it truly means, and how good it is. In parallel, you must clean the data. When I was at Bell Labs in the 1980s and 1990s, we used the expression “rinse, wash, scrub” for increasingly sophisticated efforts to find and correct errors (or at least eliminate them from further analyses). For big data, a complete rinse, wash, and scrub may prove infeasible. An alternative is to complete the rinse, wash, scrub cycle for a small sample, repeat critical analyses using this “validated” data, and compare results. To be clear, this alternative must be used with extreme caution!

But simply cleaning up erred data is not enough. The sheer quantity of new data being created or coming in is growing too rapidly to keep up. Over the long term, the only way to deal with data quality problems is to prevent them.

Prevent the problems that haven't happened yet. Here is where scientific traditions of “getting close to the data” and “building quality in” are most instructive for big data practitioners. I've already mentioned the care scientists take to design their experiments, the efforts they make to define terms, and the lengths they go to in order to understand end-to-end data collection. They also build controls (such as calibrating test equipment) into data collection, identify and eliminate the root causes of error, and upgrade equipment every chance they get. They keep error logs and subject their data to the scrutiny of their peers. This list can go on and on.

Those pursuing big data must adapt these traditions to their circumstances. Most really important data is used for many things (not just big data analyses), so you must specify the different needs of people who use it. Since the data originates from many sources, you must assign managers to cross-functional processes and to important external suppliers and ensure that data creators understand what is expected. You must measure quality, build in controls that stop errors in their tracks, and apply Six Sigma and other methods to get at root causes. You must recognize that everyone touches data and can impact quality, so you must engage them in the effort.

Interestingly, once you get the hang of it, none of the work to prevent errors is particularly difficult. But too many organizations don't muster the effort. There are dozens of reasons — excuses, really — from the belief that “if it is in the computer, it must be the responsibility of IT” to a lack of communication across silos to blind acceptance of the status quo. While I don't want to minimize these issues, none stands up to scrutiny.

As I've opined elsewhere, it is time for senior leaders to get very edgy about data quality, get the managerial accountabilities right, and demand improvement. For bad data doesn't bedevil just big data. It fouls up everything it touches, adding costs to operations, angering customers, and making it more difficult to make good decisions. The symptoms are sometime acute, but the underlying problem is chronic. It demands an urgent and comprehensive response. Especially from those hoping to succeed with big data.

FEATURED COMMENT FROM HBR.ORG

Always interesting, thanks for the insights on the “how to” of cleaning data. —Lbroome

WHAT IF GOOGLE HAD A HEDGE FUND?

BY MICHAEL SCHRAGE

Google's Sergey Brin knows a multibillion investment opportunity when he sees one. Acutely aware of the competitive edges timely data offers sophisticated investors, the company's ever-entrepreneurial cofounder once proposed that Google launch a hedge fund. After all, what company on Earth enjoyed greater access to more insights more quickly from more people searching for more information? Then Google CEO Eric Schmidt was appalled. "Sergey, among your many ideas, this is the worst." (The legal and regulatory entanglements were apparently deemed too daunting to confront.) What a pity.

But Brin's commercial instincts were inarguably correct. The world's biggest and fastest search engine can't help but generate terabytes and petabytes of actionable investment intelligence. The company might well have given Temasek, Pimco, and Berkshire Hathaway a run for their money. In fact, Google chief economist Hal Varian (and other economists and forecasters) already use the search engine's data and analytics to predict economic futures. Google Trends is fast becoming an indispensable tool for many analysts who believe — as Brin does — that the wisdom (and/or delusions) of Googling crowds can successfully inform investment strategies worldwide. Google may not have a hedge fund, but it's unlikely that high-IQ hedge funds aren't using Google's data to better manage their own situational awareness and risk.

Google, of course, is hardly the only digital juggernaut supersaturated in an embarrassment of informational riches. What kind of hedge funds might an Amazon set up (remember, Jeff Bezos actually came from a New York hedge fund)? With the company's retail insights into the browsing, shopping, and purchasing behaviors of its customers, its views into entrepreneurial investment through its Web Services offering, and its Kindle and Fire media consumption devices, Amazon is extraordinarily well positioned to strategically invest in explicit companies or particular sectors. The same investment logic holds for Apple's innovation ecosystem; the flow and fortune of its third-party apps development alone would yield valuable insight.

Any data-driven, analytics-obsessed trader or investor would be thrilled running a Facebook fund. LinkedIn, Twitter, and other social media platforms couldn't help but induce precisely the sorts of proprietary insights that might make even a Warren Buffett sit up and take notice. These novel data opportunities indeed reflect "fundamental value." That's as true for a Walmart or a JCPenney as it is for a McKinsey & Co. or a PricewaterhouseCoopers. Virtually all growing businesses are becoming more virtual. The proliferation of digital media means that all organizations are becoming data-rich. The rising ability to identify, capture, and repurpose the data by-products of an ongoing business is coming to rival the perceived "core competence" of the core business itself.

So what's the best way for executives to rethink their relationship with the gigs, teras, petas, and exas they'll be accessing over the next few years?

The classic business perspective typically views these data as high-performance fuel to make the business run faster, better, and leaner. That's the essence of the big data business case. But that mind-set's simply too focused and operational. Truly strategic leaders — and their boards — should treat Brin's "worst idea" as their most provocative thought experiment for revisiting how they recognize and value data.

Every CEO, CMO, CFO, and business unit leader should ask themselves these questions: if I decided to launch a hedge fund based solely on the data our business generates, what kind of financial opportunities immediately suggest themselves? Who would want to invest with us? Would we do better as traders or investors? Would our opportunities be more global? Does our data suggest exciting investment options in adjacent or complementary industries? How should we be analyzing our data differently? If we collected or connected slightly different data to what we have now, could our fund do even better?

Warren Buffett is famous for observing "I am a better investor because I am a businessman and a better businessman because I am an investor." The data-driven "hedge fund hypothetical" challenges the business to reformulate its value proposition in the context of orders of magnitude of more data. Conversely, figuring out how to maximize ROI on terabytes and gigabytes can't help but force leadership to reevaluate how orders of magnitude of more data should make their business more valuable. The more data-centric or Web 2.0ish an organization becomes, the more quickly and easily portfolios of interactional, behavioral, and transactional opportunities emerge. Marry the Buffett and the Brin to get a glimpse of an alternate future.

Yes, this exercise will surface all manner of ethical — and possibly legal — conflicts and risks. But those are exactly the kinds of values conversations (privacy, exclusivity, fairness, and intellectual property) innovative organizations need to have. The hedge fund hypothetical is designed to be cognitively uncomfortable.

If it feels a little too contrived and uncomfortable, take a simpler first step: picture the kind of hedge fund your most serious competitor would run.

GET STARTED WITH BIG DATA: TIE STRATEGY TO PERFORMANCE

BY DOMINIC BARTON AND DAVID COURT

Large-scale data gathering and analytics are quickly becoming a new frontier of competitive differentiation. In a recent *Harvard Business Review* article we explore how companies require three mutually supportive capabilities to fully exploit data and analytics: an ability to identify and manage multiple sources of data, the capacity to build advanced analytic models, and the critical management muscle to transform the organization.

Getting started on a successful data and analytics journey, however, is a continuing challenge for many leaders, and they often struggle with a clear strategy that ties data and analytics to improved performance. We took a close look at companies that have recently launched big data strategies to shed further light on the tough road C-level executives face. From these experiences, we have distilled four principles for defining a strategy and getting started.

1. Size the opportunities and threats

Opportunities may range from improving core operations to creating new lines of business — even in the same industry. For example, insurance companies can use big data to improve underwriting performance now, while over the longer term they can use it to serve formerly unprofitable customers and ultimately even develop entirely new risk-based businesses. The key is to establish a clear-eyed view of the business impact expected at each stage of implementation in order to better focus efforts and determine priorities.

In the case of a retailer we studied, data and analytics were part of a difficult battle for market share. The company's strategy had long been predicated on matching the moves of an efficient big box rival, yet now a different online player was draining the company's revenues and denting its margins. At the heart of the threat was the new competitor's ability to gather and analyze consumer data to generate recommendations across millions of customers while becoming a platform where vendors could sell excess inventory at a discount by using publicly available price data. Responding to this threat required both debate on "what business are we in?" and investment to use data and analytics to drive important performance improvements.

2. Identify big data resources ... and gaps

Framing the basics of a big data strategy naturally leads to discussions about the kinds of information and capabilities required. For example, a review will have to consider access to analytical talent as well as potential partnerships that might help fill gaps. We often find that consideration of required internal and external data will often spark "aha" moments — as executives identify "data gems" cloistered inside their business units or recognize the value of cre-

ating the right kind of partnership.

The retailer mentioned above found that the company gathered volumes of data but wasn't using it to potential. This information on product returns, warranties, and customer complaints contained a wealth of information on consumer habits and preferences. The review also revealed that none of the information was integrated with customer identification data or sufficiently standardized to share within or outside the company. Happily, the company had a team that could help solve these problems: in-house data analysts whose siloed efforts were underused.

3. Align on strategic choices

Once companies identify an opportunity and the resources needed to capitalize on it, many rush immediately into action-planning mode. This is a mistake. Data strategies are likely to be deeply intertwined with overall strategy and therefore require thoughtful planning when a company decides how its resources should be concentrated to achieve the desired results.

It's also important to view data and analytics in the context of competing strategic priorities. In the case of a telecom provider, a cross-functional executive committee was created to oversee the analytics team and to ensure that its efforts were aligned with the company's strategy. The committee focused the team's efforts on two questions: "How competitive are our brands in the minds of users when they make purchase decisions?" and "What key buying factors matter for users, and how well positioned are we to communicate with customers about these factors?"

The team then combined customer data from several sources to surface actionable insights — for instance, sports and other premium TV programming was a key differentiator in purchasing decisions, and customers would be more inclined to purchase a "triple play" service offering (television, high-speed Internet, and voice telephony) if the company de-emphasized voice telephony in its marketing messages. This was the opposite of what consumers had indicated in traditional market research interviews. The analysis also underscored — and helped quantify for executives — the importance of a bigger strategic imperative: the need to add mobile telephony as a fourth service to complete a "quadruple play."

4. Understand the organizational implications

Finally, it's important to note that the threats and opportunities associated with big data often have organizational implications that only concerted senior executive attention can address. For example, at another telecom player, the consumer data insights team learned that two things led to the most rapid spread of nega-

tive word of mouth about the company on social media and micro-blogging sites: network outages and any perception by customers that the company had made false advertising claims about its products or network. Initially the marketing and network organizations, rather than cooperating, blamed one another for the findings. Only when senior executives forced the two sides to work more closely together and build trust could the company capitalize on the information by tailoring marketing messages to better explain new prod-

uct rollouts and network upgrades.

Finally, we often see stresses on technical and analytical resources as a company seeks to capitalize on data and analytics. Thus, whether a company is planning a single, large initiative or multiple smaller ones, its senior team must stay mindful of the resources required (technological and otherwise) to shift quickly from pilot to “at scale” implementation.

11:00 AM OCTOBER 2, 2012

BIG DATA'S BIGGEST OBSTACLES

BY ALEX “SANDY” PENTLAND

In my last blog about big data, I offered a very optimistic view of its promise: big data can allow us to see and predict human behavior objectively. What starts small — for instance, seeing through data how people really eat and move — can become massive, such as overhauling the health care system to address real behavior.

I am optimistic about big data, but I'm also realistic. There are many obstacles to getting to a good place with it. Here are some of them.

The Correlation Problem: When your volume of data is massive, virtually any problem you tackle will generate a wealth of “statistically significant” answers. Correlations abound with big data, but inevitably most of these are not useful connections. For instance, your big data set may tell you that on Mondays people who drive to work rather than take public transportation are more likely to get the flu. Sounds interesting, and traditional research methods show that it's factually true. Jackpot!

But why is it true? Is it causal? Is it just an accident? You don't know. This means, strangely, that the scientific method *as we normally use it* no longer works, because there are so many possible relationships to consider that many are bound to be statistically significant. As a consequence, the standard laboratory-based question-and-answer process — the method that we have used to build systems for centuries — begins to fall apart.

What we have to come up with are new ways to test the causality of connections in the real world far more frequently and earlier than we have ever had to do before. We can no longer rely on laboratory experiments; we need to actually do the experiments in the real world.

This will be disconcerting to many. We live in an era that builds on centuries of science, and our methods of building systems, governments, organizations, and so on are well defined. But with the coming of big data, we are going to be operating very much out of our old, familiar ballparks.

The “Human Understanding” Problem. Finding correlations in data is one thing. Understanding them in a way that allows you to build a new, better system is much more difficult. There needs to be

a dialogue between our human intuition and the big data statistics, and that's not something that's built into most of our management systems today. Take the flu example. How do we act on that? Do we believe it? What does our intuition about such a fact tell us to do? Managers have little concept of how to use big data analytics, what they mean, and what to believe.

In fact, the data scientists themselves don't have much intuition either, and that's an even bigger problem. One estimate recently suggested that 70 to 80 percent of the results that are found in the machine learning literature — which is a key big data scientific field — are probably wrong, because the researchers didn't understand that they were overfitting the data. They didn't have that dialogue between intuition and causal processes that generated the data on the one hand and the statistics on the other hand. They just fit the model and got a good number and published it (and the reviewers didn't catch it, either). That puts bad data out there in the world where it's acted on by practitioners who likewise don't have the understanding of the data to act critically and appropriately.

If we start building our world on results like these, we're going to end up with disastrous results.

The Provenance Problem. Earlier this year, I ran a big data session at Davos and heard from the CEOs of leading companies providing services in this area. They said that the biggest problem they faced in getting started on a big data application was getting the data out of silos and into a form where it could be used.

But this isn't just a garden variety, get-your-departments-sharing kind of corporate problem. It's more difficult than that, because with big data it is typical that no one company owns all the data you need; you need new types of collaboration, both with your customers and with other companies that serve your customers.

How do you get the data out of those silos? The first step is to figure out who owns that data, which isn't always clear. Does the telephone company own information about your location while you were on the phone? Maybe they have some right to it. What if the data is attendant to a transaction with a merchant? Who con-

trols that? Who can use and reuse that data? (And you thought the Telecom Act was complex.)

Unfortunately for most of the people in the room at Davos, this was a brand-new concept, and they weren't up to speed on it at all.

The Privacy Problem. Just as businesses are beginning to see the power of big data, consumers are beginning to ask about their right to prevent the collection and use of every bit of data they leave behind. You can imagine using big data to make a world that is incredibly invasive, incredibly "Big Brother" . . . George Orwell's 1984 vision pales in comparison.

For the last several years, I've been helping run sessions at the World Economic Forum around sourcing personal data and ownership of the data, and this effort has ended pretty successfully with what I call the New Deal on Data. The chairman of the Federal Trade Commission, who's been part of the group, put forward the

U.S. Consumer Data Privacy Bill of Rights (PDF), and in the EU the justice commissioner declared a version of this New Deal to be a basic human right.

Both of these regulatory declarations put the individual much more in charge of data that's about them. This is a major step in making big data safer and more transparent as well as more liquid and available, because people can now choose to share data. It's a vast improvement over having the data locked away in industry silos where nobody even knows it's there or what's being done with it.

FEATURED COMMENT FROM HBR.ORG

Great post! I particularly like the fact that [Sandy] brought up the "overfitting" issue. Beyond just technology, this is a cultural issue scientists and data heroes have to get past. —Bruno Aziza

7:00 AM OCTOBER 3, 2012

TO SUCCEED WITH BIG DATA, START SMALL

BY BILL FRANKS

While it isn't hard to argue the value of analyzing big data, it is intimidating to figure out what to do first. There are many unknowns when working with data that your organization has never used before — the streams of unstructured information from the Web, for example. Which elements of the data hold value? What are the most important metrics the data can generate? What quality issues exist? As a result of these unknowns, the costs and time required to achieve success can be hard to estimate.

As an organization gains experience with specific types of data, certain issues will fade, but there will always be another new data source with the same unknowns waiting in the wings. The key to success is to start small. It's a lower-risk way to see what big data can do for your firm and to test your firm's readiness to use it.

The Traditional Way

In most organizations, big data projects get their start when an executive becomes convinced that the company is missing out on opportunities in data. Perhaps it's the CMO looking to glean new insight into customer behavior from Web data, for example. That conviction leads to an exhaustive and time-consuming process by which the CMO's team might work with the CIO's team to specify and scope the precise insights to be pursued and the associated analytics to get them.

Next, the organization launches a major IT project. The CIO's team designs and implements complex processes to capture all the raw Web data needed and to transform it into usable (structured)

information that can then be analyzed.

Once analytic professionals start using the data, they'll find problems with the approach. This triggers another iteration of the IT project. Repeat a few times, and everyone will be pulling their hair out and questioning why they ever decided to try to analyze the Web data in the first place. This is a scenario I have seen play out many times in many organizations.

A Better Approach

The process I just described doesn't work for big data initiatives because it's designed for cases where all the facts are known, all the risks are identified, and all steps are clear — exactly what you won't find with a big data initiative. After all, you're applying a new data source to new problems in a new way.

Again, my best advice is to start small. First, define a few relatively simple analytics that won't take much time or data to run. For example, an online retailer might start by identifying what products each customer viewed so that the company can send a follow-up offer if the person doesn't purchase. A few intuitive examples like this allow the organization to see what the data can do. More important, this approach yields results that are easy to test to see what type of lift the analytics provide.

Next, instead of setting up formal processes to capture, process, and analyze all the data all the time, capture some of the data in a one-off fashion. Perhaps a month's worth for one division for a certain subset of products. If you capture only the data you need to

perform the test, you'll find the initial data volume easier to manage, and you won't muddy the water with a bunch of other data — a problem that plagues many big data initiatives.

At this point, it is time to turn analytic professionals loose on the data. Remember: they're used to dealing with raw data in an unfriendly format. They can zero in on what they need and ignore the rest. They can create test and control groups to whom they can send the follow-up offers, and then they can help analyze the results. During this process, they'll also learn an awful lot about the data and how to make use of it. This kind of targeted prototyping is invaluable when it comes to identifying trouble and firming up a broader effort.

Successful prototypes also make it far easier to get the support

required for the larger effort. Best of all, the full effort will now be less risky because the data is better understood and the value is already partially proven. It's also worthwhile to learn that the initial analytics aren't as valuable as hoped. It tells you to focus effort elsewhere before you've wasted many months and a lot of money.

Pursuing big data with small, targeted steps can actually be the fastest, least expensive, and most effective way to go. It enables an organization to prove there's value in a major investment before making it and to understand better how to make a big data program pay off for the long term.

FEATURED COMMENT FROM HBR.ORG

Bill, thank you for writing this — someone had to say it.

—Anonymous

4:39 PM OCTOBER 3, 2012

THE APPLE MAPS DEBATE AND THE REAL FUTURE OF MAPPING

BY DENNIS CROWLEY

The news of the past couple of weeks about the stark differences between Apple's and Google's maps has shed light on how hard it is to build a mobile map. Showing a destination that's a few hundred yards off becomes a critical flaw, and there are tens of millions of such place markers on these maps. While the narrative about Apple's recent maps release has largely focused on the base utility of maps — that is, navigation — the reality is that building the map of the future is something that's even more ambitious and difficult.

The future of mobile maps is not just satellite imagery and a drag-able interface (as amazing as those things are). It's about access to real-time information about everything around you, from deals to popular places to where your friends are. In short, the future of mapping is the social map. (Full disclosure: I'm the CEO of a social mapping company.) In an age where so much technology feels magical, the map should be no exception. We should be building something enchanting, like Harry Potter's Marauder's Map, the mystical document that showed Harry and his friends exactly what was going on at Hogwarts School of Witchcraft and Wizardry in real time — including every corner of the castle (even secret corridors) — as well as where any given person was located in real time. (Of course, even Hogwarts should have had opt-in privacy policies so people could choose whether or not to disclose their locations.)

We should be building maps that not only orient you to your physical surroundings but also help you discover things about your surroundings, such as where to find the best mozzarella sticks or where your friends are hanging out at this very moment or an alert that the shoe store you just walked by is having a sale on your favorite brand. A map as a navigation tool is relatively static and

barren. Maps that open to a blank page are boring. Social maps, on the other hand, are active. They're heavily populated with dots and icons that represent points and people of interest around you. The technical requirements to deliver a map like this are profound but possible to pull off by leveraging three current phenomena: crowdsourcing, big data, and the social graph.

Crowdsourcing, where a community of users populates an ecosystem with content, is the best way to create an expansive and accurate map. Think of Wikipedia for cartography. In the same way that you might go to Wikipedia for explanations of historical moments written by people who are most familiar with the subject matter, on a map you want business owners or their most ardent supporters creating their digital profiles. Community members become stewards of their portion of the content because it's near and dear to them. Crowdsourced information can then be algorithmically ranked based on how it's used by the rest of the community, so the best and most accurate data rises to the top.

Once you have a trusted and secure base layer of location-based data, you can do lots of really interesting things on top of it. You can start to take these billions of pieces of user-generated data and deliver highly pertinent information, all in milliseconds. *This* is the opportunity that big data presents. The season, time of day, and current weather all can act as valuable signals to the recommendation engine that decides who receives what content and when. Searching for something to do on a rainy day? Nearby museums should rank higher than parks in search results. Community activity is another important data set. By tapping the community at large, users can see which businesses are busy and which aren't.

And then there's personalization, which is where things get *really* interesting, because relevance is relative . . . to you. It's relative to your likes and dislikes, your previous behavior, and what you're in the mood for. Sometimes you know exactly what you're looking for, but what about when you're feeling less decisive? You know you want Italian food, but you're in an unfamiliar neighborhood. How do you distinguish between the options that a traditional mobile map yields? Wouldn't it be great if the Italian place that most closely resembles your favorite spot back home rises to the top of the list? Local search should no longer be one-size-fits-all.

And that brings us to the social graph, the intricate diagram of your connections to people and things via the Web. Let's say you're on a trip to Albuquerque and you're looking for a place to stay. You search your phone, and rather than giving you a generic, directory-style list of hotels and motels around you, it produces a ranking of places to stay based on the previous behaviors of you and your friends. Even if no one has been to Albuquerque, it can draw

comparisons between the local places and the places you and your friends frequent all over the world. Or what if you're at the front desk of your newly discovered B&B and the concierge offers you room options? You check your phone to see whether any of your friends have stayed here, and voilà — your friend has left a tip saying it's well worth an extra \$20 to pay for the suite. That's the real-world version of the Marauder's Map showing you the secret path.

So the discussion over the base layer of mobile maps is an important one, but the future of mapping lies in the wonders that can be built on top of these platforms — the social map. Those innovations are changing modern-day exploration. The quickest way from point A to B will always be a straight line (and, of course, we first have to make sure that we can reliably get people from A to B), but what are the points of interest along that path and where are the good times happening should you decide to take a detour? Those are the questions we should *really* be asking.

12:00 PM OCTOBER 4, 2012

WHY DATA WILL NEVER REPLACE THINKING

BY JUSTIN FOX

Big data, it has been said, is making science obsolete. No longer do we need theories of genetics or linguistics or sociology, *Wired* editor Chris Anderson wrote in a manifesto four years ago: "With enough data, the numbers speak for themselves."

Last year, at the Technomy conference outside Tucson, I heard Vivek Ranadivé — founder and CEO of financial data software provider TIBCO, subject of a Malcolm Gladwell article on how to win at girls' basketball, and part owner of the Golden State Warriors — say pretty much the same thing: "I believe that math is trumping science. What I mean by that is you don't really have to know why, you just have to know that if A and B happen, C will happen."

Anderson and Ranadivé are reacting to something real. If the scientific method is to observe, hypothesize, test, and analyze, the explosion of available data and computing power has made observation, testing, and analysis so cheap and easy in many fields that one can test far more hypotheses than was previously possible. Quick-and-dirty online "A/B tests," in which companies like Google and Amazon show different offers or page layouts to different people and simply go with the approach that gets the best response, are becoming an established way of doing business.

But that does that really mean there are no hypotheses involved? At Technomy, Ranadivé made his math-is-trumping-science comments after recommending that the Federal Open Market Committee, which sets monetary policy in the United States, be replaced

with a computer program. Said he, "The fact is, you can look at information in real time, and you can make minute adjustments, and you can build a closed-loop system, where you continuously change and adjust, and you make no mistakes, because you're picking up signals all the time, and you can adjust."

As best I can tell, there are three hypotheses inherent in this replace-the-Fed-with-algorithms plan. The first is that you can build U.S. monetary policy into a closed-loop system; the second is that past correlations in economic and financial data can usually be counted on to hold up in the future; and the third is that when they don't, you'll always be able to make adjustments as new information becomes available.

These feel like pretty dubious hypotheses to me, similar to the naive assumptions of financial modelers at ratings agencies and elsewhere that helped bring on the financial crisis of 2007 and 2008. (To be fair, Ranadivé is a bit more nuanced about this stuff in print.) But the bigger point is that they are hypotheses. And because they'd probably prove awfully expensive to test, they'll presumably stay hypotheses for a while.

There are echoes here of a centuries-old debate, unleashed in the 1600s by protoscientist Sir Francis Bacon, over whether deduction from first principles or induction from observed reality is the best way to get at truth. In the 1930s, philosopher Karl Popper proposed a synthesis in which the only scientific approach was to formulate

hypotheses (using deduction, induction, or both) that were falsifiable. That is, they generated predictions that — if they failed to pan out — disproved the hypothesis.

Actual scientific practice is more complicated than that. But the element of hypothesis/prediction remains important, not just to science but also to the pursuit of knowledge in general. We humans are quite capable of coming up with stories to explain just about anything after the fact. It's only by trying to come up with our stories beforehand, then testing them, that we can reliably learn the lessons of our experiences — and our data. In the big data era, those hypotheses can often be bare bones and fleeting, but they're still always there, whether we acknowledge them or not.

“The numbers have no way of speaking for themselves,” political forecaster Nate Silver writes in response to Chris Anderson near the beginning of his wonderful new doorstopper of a book *The Signal and the Noise: Why So Many Predictions Fail — But Some Don't*. “We speak for them.” He continues, “Data-driven predictions can succeed — and they can fail. It is when we deny our role in the process that the odds of failure rise. Before we demand more of our

data, we need to demand more of ourselves.”

One key role we play in the process is choosing which data to look at. That this choice is often made for us by what happens to be easiest to measure doesn't make it any less consequential, as Samuel Arbesman wrote in Sunday's *Boston Globe*: “Throughout history, in one field after another, science has made huge progress in precisely the areas where we can measure things — and lagged where we can't.”

In his book, Silver spends a lot of time on another crucial element, how we go about revising our views as new data comes in. Silver is a big believer in the Bayesian approach to probability, in which we all have our own subjective ideas about how things are going to pan out but follow the same straightforward rules in revising those assessments as we get new information. It's a process that uses data to refine our thinking. But it doesn't work without some thinking first.

FEATURED COMMENT FROM HBR.ORG

Data is an important ingredient of making informed decisions, but critical thinking is what actually makes it relevant. —Oscar Trelles

11:00 AM OCTOBER 5, 2012

BIG DATA DOESN'T WORK IF YOU IGNORE THE SMALL THINGS THAT MATTER

BY ROBERT PLANT

Ever waited hours in vain for a repair service to arrive at your home? Of course you have. We all have. Chances are you've also shifted your allegiance away from a company that made you wait like that.

So why do companies spend millions on big data and big data-based market research while continuing to ignore the simple things that make customers happy? Why do they buy huge proprietary databases yet fail to use plain old scheduling software to tell you precisely when a technician is going to arrive?

For that matter, why do they send trucks with the wrong inventory on board? Why do they impose unfathomable password requirements? Create complex online registration processes? Shut down their customer service on weekends? Hound you with robo-calls until you answer a survey?

Big data is today's panacea, the great new hope for unlocking the mysteries of marketing. To avoid being left behind, companies are rushing to cash in on the information they glean from customers, and vendors are stepping up to help. The venture capital community has taken note of this trend: over the past year, investors have poured hundreds of millions of dollars into start-ups that promise to exploit caches such as Facebook's reported 100 petabytes of data.

But in the meantime, most companies haven't done much to improve the customer experience. Negative interactions with com-

panies are as common as ever. A reported 86 percent of consumers have switched companies after bad experiences as customers.

Companies would do better at satisfying and retaining customers if they spent less time worrying about big data and more time making good use of “small data” — already available information from simple technology solutions — to become more flexible, informative, and helpful.

Miami-Dade Transit allows customers to monitor its system so they can know when to arrive at a stop in time to catch a particular bus. That's not so complicated; why can't commercial firms invest in allowing customers to monitor the location of service or delivery techs? The U.K. grocery chain Tesco gives customers a choice of delivery times that come with different prices (deliveries are cheaper during off-peak hours). That's not complicated either; why do so many companies give you no choice at all about deliveries? Cemex, the concrete maker, uses scheduling software to give customers a 20-minute window for deliveries. Comcast promises a \$20 credit or three months of a free premium channel if techs are late. After a recent credit card mix-up, Apple responded immediately to my e-mail and solved my problem. Why do so many other companies mishandle customer service and leave you to fend for yourself? Many innovative customer service solutions such as these are

simple and inexpensive, and they have high payback.

Big data, by contrast, is far from inexpensive, and the payback is often iffy. Last year the U.K. government pulled the plug on an expensive and unwieldy patient record database that was to have been used for clinical research and improved business processes. The government shifted its focus to creating smaller projects on more well-defined deliverables such as on-time patient appointments.

CXOs with long memories may recall prior panaceas that underdelivered over the long run. Think of “expert systems,” which promised to deliver expert levels of performance but were unable to provide explanations for why the recommended courses of action should be taken, a failing that limited their usefulness. (Expert systems are becoming fashionable again through IBM’s Watson.)

I’m not saying that all big data projects are useless. Far from it. Manchester City Football Club, the English Premier League champion, has opened up part of its “on-ball events” database in the hope that people in the open data community will find patterns and trends that could give the club an edge. That’s a simple way to make use of a big trove of data, and the upside could be substantial.

But don’t expect an easy payoff. To avoid being caught spend-

ing vast sums on half-vast results, CXOs would be wise to link a big data project to the development of a rigorous metrics program — something like the Balanced Scorecard, which is more likely than CapEx or financial ROI to capture the full results of big-data applications downstream and across multiple process groups. Let’s say, for example, that you’re seeing a pattern of strong store sales for a group of products that were previously perceived as unrelated. It might take a scorecard approach for you to figure out that the sales peak coincided with a particular phase in the staff training schedule. A scorecard that links financials with learning initiatives and other operations would serve as a cross-check for managers.

And in the frenzy to capitalize on big data, don’t forget what it’s like to be a data *point* — an individual customer dealing with your company. If you’re not making your data points happy, they’ll gladly move into someone else’s database, just as you did after the repair service failed to show up.

FEATURED COMMENT FROM HBR.ORG

The thrust of Dr. Plant’s article is spot on — be excellent at your operations with the data you have before looking to capture and analyze more. —Cryptoideologist

2:55 PM OCTOBER 9, 2012

WHAT SHOULD YOU TELL CUSTOMERS ABOUT HOW YOU’RE USING DATA?

BY NIKO KARVOUNIS

If you’re a senior manager launching a big-data initiative, you should start by asking three simple, high-level questions to guide your organization’s data collection strategy. Once you have an analytics strategy in place, it’s time to think about how you’re going to apply the data you’re collecting in the marketplace. How will you use it in interactions with customers, competitors, and collaborators? It’s especially important to anticipate how your use of data will be perceived by your customers. Here are some crucial questions to consider when it’s time to start feeding data back into the marketplace.

How transparent will we be about the data we use to make decisions? Clearly, businesses using sensitive data should be up-front and open about what they’re collecting and how it’s used. But there’s a second level of transparency that might be better thought of as obviousness: how blatantly do you want to link your external activities to your data?

Consider the case of Target. The retail giant tracks customer behavior so rigorously that it had accurately known when a teen-aged customer was pregnant based on her buying patterns; this was before she had even told her father, who was outraged when Target started sending her baby product promotions. (It turns out

consumers aren’t comfortable with the idea that a big box retailer knows them better than their own family does!) Target didn’t stop collecting and leveraging customer behavior data in its marketing, however. Instead, it simply became more subtle in its execution: customer-specific offers are now interspersed with more generic ads or a more diverse set of promotions.

Sometimes data-savvy companies get so excited by their analytical horsepower that they don’t stop to think of how their business intelligence is perceived by outside parties. This is particularly relevant to marketing, where customer experience is key. However, it’s also relevant when you think about what signals your activities may send to competitors. If your big data capabilities are a sensitive competitive advantage, delicacy in communicating what and how much you know may be important. Subtlety sometimes has its place.

How accessible and usable do we want our data products to be? Today data can be integrated directly into products in ways never before possible. Think about Nike+, which allows runners to track and share their performance, or Weight Watcher’s eTools, which provide an online dashboard for dieters. But translating data and analytics into a product offering raises a number of questions. How

much data do you want to share with your customers — just theirs or data from others as well (e.g., comparing running times)? How specific do you want the data to be (e.g., direct comparisons or just rankings)? How can you make potentially complicated data more digestible and intuitive (e.g., through scoring systems or data visualizations)? How broadly accessible is your data — do users have to register, pay, or meet certain criteria to engage with your analytics? At Quovo we've learned quickly that when data becomes a key part of the customer experience, it's important to think like a product manager as well as a data strategist. The more comfortable consumers become with data, the truer this will become in the future.

Are there win-win partnerships that will help us be smarter about our data? As I suggested earlier, good data is often viewed as a competitive advantage to protect fiercely; but as information becomes more ubiquitous, a different, more collaborative mind-set can be beneficial.

Consider: as the world's largest online social network, Facebook already has a trove of data on its users. But through its Facebook

Connect program, which allows users on Facebook to sign onto other sites using their Facebook credentials, the network can also track behaviors such as "Likes" on external sites. In return, partner sites can offer visitors an easier log-in to access and share content through Facebook's newsfeed, which should help increase traffic.

As the big data landscape evolves, similar opportunities to scale data through collaboration will become more common, including through third-party vendors. Organizations such as Metamarket and LexisNexis Risk Solutions (full disclaimer: I've worked for LNRS) manage contributory databases that pool data from multiple companies. In return for contributing, each organization gets access to data and analysis that would have been beyond its reach had it limited itself to only its own data set. Not every company is going to be comfortable with this arrangement, and data privacy is an important consideration; but as data volume grows across industries, businesses will soon need to find ways to scale up their data insights quickly. Partnerships may be a good way to get smarter faster.

1:00 PM OCTOBER 10, 2012

DATA CAN'T BEAT A SALESPERSON'S BEST TOOL

BY RICK REYNOLDS

For people selling in a B2B context, big data can be extremely useful — but amid all the celebration over this powerful new way of doing business, it's important to keep in mind its limits. There is an important piece of information missing from most B2B big data collections: firsthand qualitative data about what your prospects and existing customers really think about your company. In other words, beyond the data, what do your customers say when you are not around?

As sales becomes more technologically advanced, the trend is for B2B companies to rely more on big data and less on information gathered from one-on-one personal interactions. The big data that enterprise companies are collecting is often limited to numbers of sales transactions, competitive data, sales activity data, and overall market trends. All these facts and figures are relevant and useful for improving sales, but the data alone doesn't provide in-depth insight into what is really happening in the minds of your prospects and clients. Much as in the intelligence world, satellites can provide only so much insight. Some of the most critical data can be obtained only through one-on-one, face-to-face interactions.

At one extreme, many companies don't have big data sets. They rely on sales teams to collect and manage whatever data they are able to obtain through their own research. In fact, according to the CSO Insights 2012 Impact of Data Access on Sales Performance

Report, the average sales rep spends 24 percent of his or her time searching for relevant information to prepare for calls. This, though, takes away a significant portion of selling time.

At the other extreme, many enterprise companies have so much big data that it is difficult to interpret and manage, let alone use effectively to increase bottom-line results. The same study by CSO also found that nearly 90 percent of sales leaders report missing opportunities due to information overload. Sifting through big sales data can be like trying to pick out a molecule in a raging river.

Quantitative big data about prospects and existing clients is critical to sales success, but it does not provide a complete picture. Your company needs both quantitative and qualitative data to fully understand your customers. In other words, you need the facts, figures, and the context.

Also keep in mind that your big data may mislead you at times as it relates to individual clients. Not every trend applies to all companies. In fact, your most valuable multimillion-dollar client may buck a trend altogether. If you assume the trend applies to this client, you risk losing that customer. Let's take a look at what kind of qualitative data you need to better understand your customers and prospects. You need to know their:

- Strategic direction
- Stated needs

- Hidden needs not mentioned in an RFP
- Decision-making process and parties involved
- Perception of your performance and customer service
- Perception of your competitors

There is only one way to uncover qualitative data about your B2B prospects and existing clients — ask for it. This level of in-depth information can be obtained only by asking probing questions, carefully listening to the responses, and analyzing what you uncover. However, it requires more than developing a brief survey. To get to this level of understanding, in-depth 20- to 60-minute interviews should be completed with executives and managers at your client and prospect companies.

When conducting in-depth interviews, be sure to interview multiple executives at the senior level, as their views truly drive company direction. Ask them a structured series of probing, open-ended questions to get at the truth. You need to scratch below the surface. If you think your prospects or clients won't provide you with candid feedback, or if you feel you lack the required exper-

tise, bring in an objective third party to conduct the interviews and analysis.

Collecting qualitative data is only half the process. The other half is acting upon what you have learned. Once your prospects and clients confide in you, they expect you to act on what you have uncovered. Failure to make improvements and changes based on what you learned will jeopardize the account and set you back.

Uncovering both quantitative and qualitative insights about prospects and existing clients provides a comprehensive picture to help you more effectively meet your customers' needs and improve your performance. One type of data without the other provides only half the picture. However, when combined, quantitative and qualitative data about your prospects and customers will help you improve sales and better serve your customers.

FEATURED COMMENT FROM HBR.ORG

Feedback from lost clients and prospects generates excellent insights. Whether they're right or wrong, it's what they think, and that's what you need to know.

—Rick Reynolds

9:00 AM OCTOBER 12, 2012

WHEN PIRATES MEET ADVANCED ANALYTICS

BY ROBERT GRIFFIN

Crime fighters are learning to anticipate criminal behavior by finding important “clues” hidden in millions of pieces of disparate data. One fascinating example of this trend is the work that maritime professionals are doing to stop piracy.

They're already good at analyzing structured data from previous incidents and real-time oceanographic conditions — and they're definitely seeing some results. The latest statistics from the International Maritime Bureau show a 54 percent drop in reported incidents during the first half of 2012 as compared to the same period in 2011, some of which is certainly attributable to improved analytics.

But travelers and businesses still face a big challenge from pirates. U.S.-based consultancy Oceans Beyond Piracy estimates that the economic impact of piracy totaled \$7 billion in 2011.

The real game changer will be the ability to analyze unstructured data. According to Daryl Williamson, director for government business at Lloydlistintelligence.com, a U.K.-based provider of maritime information, pirates are always changing their tactics and changing course — so the trick is to uncover patterns and trends that are not immediately obvious.

There are almost too many sources of *unstructured* data to grapple with: interviews with pirates in custody, news stories about piracy incidents, data from mobile phones found during investiga-

tions, e-mail traffic, and social media posts from the pirates themselves. And here's where the story gets really interesting, in my opinion. Most of this data comes from disparate sources that can vex the best investigators. It's not simply a matter of easily formatted spreadsheets with clean rows and columns. At warp speed, data comes in from the Web, mobile devices, PDF files, and other documents — a potential treasure trove of hidden insights.

Until recently, the industry standard was to focus on bringing all the data into a proprietary, central warehouse where it would be reformatted and analyzed. That worked fine when it was simple spreadsheets and databases. Today's world is not so simple. The volume, variety, velocity, and veracity of data require a flexible analytical environment so that data can remain fluid; today's data can have a totally different meaning tomorrow. Maritime agencies need to be able to respond to new data in real or near real time.

Visualization is key. Working closely with technology partners such as Esri, the geospatial mapping experts, maritime officials are starting to create visual pictures whose primary goal is to identify, track, intercept, and disrupt a highly mobile and increasingly organized network. According to Esri, by mapping data points based on location and time details, patterns and relationships come into view. Visualization establishes nonobvious connections between

seemingly unrelated data; for example, it links persons, places, or things to a past incident or to other information about a future incident. As we've seen in law enforcement, patterns emerge that can help us get resources to the right place at the right time. We're starting to develop genuine predictive models.

I believe that advanced analytics will have a huge impact on our ability to fight crime, at sea and on land. But a key to success will be learning how to more effectively turn data into pictures so that it tells a story not just to data scientists but to security professionals as well.

9:00 AM OCTOBER 15, 2012

WHAT COULD YOU ACCOMPLISH WITH 1,000 COMPUTERS?

BY DANA ROUSMANIERE

An interview with Frederic Lalonde and Chris Lynch, serial entrepreneurs and founders of Hack/Reduce, "Boston's Big Data Hacker Space." Fred Lalonde is founder and CEO of travel start-up Hopper and former vice president at Expedia. Chris Lynch was most recently SVP and GM of HP's data analytics business unit.

HBR: You recently launched Hack/Reduce — a sort of big data playground — in collaboration with MIT, Harvard, and other Boston-area universities. What was the inspiration behind it?

Fred Lalonde: We started Hack/Reduce in response to two big barriers to using big data. One is scalability. You can't do big data on your laptop. You need 1,000 computers. The second is finding people who have the training, knowledge, and expertise to work with big data. Even in a city like Boston with its depth of software talent, there was literally nowhere where people with an idea could boot up enough computing power to try it.

We started with a one-day hack-a-thon and invited people to come over. The premise was you say to geeks, "We're booting up all the computing power you need to come and do whatever you want."

HBR: What types of experiments have people tried?

FL: We had people working on genomic sequencing to find cancer markers. People were analyzing real-time traffic patterns. A group of people determined who are the friends of U.S. congressmen and congresswomen by analyzing Twitter graphs. One group looked at best places to pick up a bike share in Montreal. (Hint: nobody returns bikes to the stations at the top of hills.) It's amazing what people will do in a day with computing power and data sets. We've seen amazing things happen in five hours. Imagine what could happen if these people had five months.

HBR: Hack/Reduce started out as a side project for you, but it sounds like it has real legs.

FL: It's now a full-blown 503c, established in partnership with the commonwealth of Massachusetts.

HBR: You're also launching a travel company, Hopper.com, that's being built from the ground up on big data. Tell us about the opportunity you saw in the travel space.

FL: All the existing travel sites work well if you know what you're looking for — a Chinese restaurant near a train station or a hotel — but planning a trip is an aspirational process, and there's a whole element of discovery that's not covered anywhere.

This month we'll crawl a billion pages about travel using our own hardware and our own data center. We're parsing this billion-plus sea of travel information and creating structure out of it. To put it in context, there are about two million travel-related blog posts created every day. There are about fifty-six million blogs on Wordpress right now, and travel is the fifth most often used tag — it's bigger than fashion and sports. This doesn't even include social networking posts. There is more superinteresting, high-quality, firsthand user experience information outside the travel sites than there is inside travel sites. But when we plan a trip today, the search engines are always returning the same user review sites. The size of the data being generated in the wild is immense, and it's just sitting out there. Part of what we're doing is trying to understand the blogosphere as a whole, finding the relevant travel data, and adding it to a catalog to make it consumable.

When we started looking at the problem, we realized that the travel catalog is broken. To use a music analogy, imagine that you didn't have a unique catalog: you heard a song but didn't know the title, and maybe you could find the artist online. If there were no structured data set for music, you could still download an MP3, but value-added services like Pandora and Spotify couldn't exist. If you look at travel, it's an equally large data set — and I'm not talking about flights and hotels, because no one travels for the flight and hotel. We travel for the destination, the experience. Imagine a complete database of every experience on Earth and every experience that's been had in that place — that's the type of travel experience you would have with big data, and that's what we set out to build with Hopper.

HBR: How do you know which data to trust?

FL: The ability to do statistical analysis — the patterns that occur at the scale of big data — is where the truth lies. We're taking a holistic view that only a group of computers can do.

HBR: What type of employee thrives in a company that's founded on big data?

FL: Almost everyone in our company has written code at some point. More generally, big data requires people who like complexity and solving big problems.

HBR: What's the potential of big data from your point of view?

CL: The potential is as disruptive as the Internet and mobile technologies have been. Big data is changing curriculums at universities, creating new roles and professions (e.g., data scientist), and

highlighting the fact that we are all content creators. The availability of the data and access to tools (in many ways thanks to the open source community) has never been greater; I anticipate the role big data analytics plays within organizations to increase dramatically over the next several years. And while big data may be overhyped, I do think it's going to solve a lot of the world's problems.

HBR: What's your advice to HBR readers who are just beginning to think about the potential of big data in their own companies?

FL: If you're an exec thinking about big data, it probably feels vague and confusing. You need to get your product people to sit down with the people who are solving data problems, and you need to start working with your own data. It's shocking what most companies are sitting on right now. The established companies are going to fall into two categories — the ones who are leveraging data and the ones who aren't. And the upside will be staggering.

WEBINAR SUMMARY | OCTOBER 4, 2012

WHAT'S THE BIG DEAL ABOUT BIG DATA?

FEATURING ANDREW MCAFEE

Andrew McAfee, Principal Research Scientist, Center for Digital Business, MIT Sloan School of Management

Angelia Herrin (Moderator), Editor for Research and Special Projects, Harvard Business Review

Overview

Analyzing data is nothing new for business leaders. But today's volume, variety, and velocity of data are like nothing seen before. Big Data, in combination with emerging technologies from numerous firms, enables companies to generate insights that weren't possible in the past.

Leaders in industries such as retail, transportation, and insurance have also embraced Big Data and the benefits it can bring to the business. Companies that profit from Big Data have data-driven cultures and value executives who use quantitative information rather than intuition for decision making.

Context

Dr. McAfee discussed how companies are using Big Data to improve their profitability and competitiveness.

Key Learnings

Big Data gives companies greater insight into customers and business operations.

Businesses have always had access to data, but that information

has often been fragmented in silos. New technologies and large volumes of data now enable companies to generate insights that weren't possible in the past.

For example, the typical supermarket gathers information related to sales, inventory, and customer purchases. Yet it has little visibility into which items customers considered but didn't buy or how the store layout affects purchases. In contrast, well-designed online stores know how consumers navigate, what products they almost purchased, and how effective promotions are.

With Big Data, this level of understanding is available to all types of businesses. Big Data is characterized by the three Vs:

- **Volume.** Today it is common for organizations to maintain petabytes of data. Looking ahead, zettabytes will be the norm.
- **Velocity.** The pace at which organizations accumulate and process information is astounding. Google handles 20 petabytes of data each day.
- **Variety.** Useful data comes from many different sources, such as mobile phones, sensors, and social media.

Mainline industries such as retail, transportation, and insurance are profiting from Big Data.

Big Data isn't for just high-tech companies. Progressive organizations in mainline industries are using Big Data and analytics to improve their bottom lines.

New database technologies efficiently deliver targeted customer promotions. Retail giant Sears is a holding company for major

brands including Lands' End and Craftsman. Tailored, cross-brand promotions represent a huge advantage for Sears, but merging customer data across the different brands has been challenging. In the past, it took eight weeks to extract information from various data warehouses, transform it to a uniform format, and load it into an analytics system. To reduce the ETL (extract, transform, load) cycle, Sears implemented Hadoop, a new database technology that facilitates proactive, on-the-fly analytics. Preparing a cross-brand promotion now takes just two weeks, and no new IT or marketing resources had to be hired.

Predictive analytics streamline business operations. Airports require accurate predictions about flight arrival times so ground crews can be ready for planes. In the past, airport staff asked pilots for arrival estimates. However, this information was frequently inaccurate, leading to waits on the tarmac and dissatisfied passengers. PASSUR Aerospace developed a database of past flight arrival times as well as information about weather and congestion. This has resulted in significantly more accurate arrival predictions for airport ground crews.

Pattern recognition leads to better health care. The American health care industry is fragmented, and data isn't usually shared across provider systems. Health care insurers, however, have access to huge amounts of patient data. Aetna analyzes information to determine whether customers are getting the best health care possible. The company knows, for example, which medications should be administered after specific procedures. If a patient doesn't receive those medications, Aetna sends a targeted message to both the patient and the doctor.

“Big Data is accelerating customer knowledge in a way that was impossible in the past.”

—Andrew McAfee

Cutting-edge Big Data projects offer a window into tomorrow's business reality.

There is no doubt that Big Data will transform the way companies do business. Innovative projects are combining new technology and large volumes of information in cutting-edge ways. Dr. McAfee described two ways that groundbreaking initiatives will influence the business world.

Solutions to data problems will come from crowdsourcing. Kaggle runs online competitions for customers that match data problems with a worldwide team of data scientists. Allstate Insurance recently gave Kaggle a large data set, and competitors were asked to predict bodily injury liability based solely on the insured vehicles' characteristics. The Claim Prediction Challenge ran for 12 weeks. During that time, the accuracy of the competitors' predictions improved 340 percent. The final predictions were 240 percent better than Allstate's information.

“With enough eyeballs, all data problems are tractable.”—Andrew McAfee

Digital assistants will be commonplace. As technology becomes more sophisticated, “digital colleagues” will help workers sift through large data sets efficiently. IBM's Jeopardy-winning Watson computer is an early example. Watson rapidly identifies patterns and relationships in millions of digital documents. These same skills could be applied to legal documents or disease-related data. Another example is Google's autonomous driving system. This car leverages massive amounts of data as it mimics and improves on human driving behavior.

“Being humble and allowing the organization to be led by data requires a managerial shift.”

—Andrew McAfee

To take advantage of Big Data, managers must adopt new decision-making methods.

As technologies advance, competition will increase and companies that see the value in Big Data will prosper. The nature of decision making will also change. Organizations will no longer rely on senior people who base decisions on intuition rather than on data.

Unlike in years past, HiPPOs (highest-paid person's opinions) will not dominate the business world. Instead, the most successful companies will be led by data-driven insights. Unfortunately, this approach is hard for many executives to adopt.

Dr. McAfee recommends four steps for leaders who want to get started with Big Data:

1. Begin with one business unit. Rather than implementing a companywide Big Data initiative, identify one business unit with data-driven managers and a fairly advanced technological environment.
2. Use the rule of five: five projects in five weeks by five people. To impose discipline on Big Data experiments, select no more than five projects. Each one should generate meaningful information within five weeks and require no more than five employees.
3. Experiment, measure, share, and replicate. Hypotheses and experiments form the core of effective Big Data projects. It is essential to measure results. Managers cannot be afraid to admit that their hypotheses are wrong, based on the data. Results should be shared as broadly as possible, and successes should be replicated using technology.
4. Remember Joy's Law. Bill Joy of Sun Microsystems said that the smartest people work for someone else. It is fine to tap into outside expertise using resources such as Kaggle to solve Big Data challenges.

Other Important Points

Focus on business problems. A common Big Data pitfall is focusing on technologies rather than on the business problems that need to be solved.

Eliminating data silos starts with dialogue. Technology alone can't prevent data fragmentation. Stakeholders must first engage in a conversation to establish the data set that represents the "single source of truth."

The information contained in this summary reflects BullsEye Resources, Inc.'s, subjective condensed summarization of the applicable conference session. There may be material errors, omissions, or inaccuracies in the reporting of the substance of the session. In no way does BullsEye Resources or Harvard Business Review assume any responsibility for any information provided or any decisions made based upon the information provided in this document.

Created for Harvard Business Review by BullsEye Resources
www.bullseyeresources.com
© 2012 Harvard Business School Publishing.

LEARN MORE

FOR MORE DETAILS ABOUT BIG DATA AND
HIGH-PERFORMANCE ANALYTICS:

→ sas.com/high-performance-analytics

A HARVARD BUSINESS REVIEW INSIGHT CENTER REPORT



hbr.org