Big Data Means Big Business

Edited by

Douglas Laney, Gartner, Inc.

Big Data Means Big Business

Edited by
Douglas Laney, Gartner, Inc.

About the Authors

Douglas Laney

Douglas Laney, research vice president, is considered a pioneer in the field of data warehousing and originated the field of infonomics (short for "information economics"). He has led analytics and information-management-related projects on five continents and in most industries. Mr. Laney is also an experienced IT industry thought leader, having launched Meta Group's Enterprise Analytics Strategies research and advisory service, established and co-led the Deloitte Analytics Institute, and guest-lectured at leading business schools on information asset management and valuation. In addition, he has led business analytics consulting and marketing practices for several software companies. Follow his blog and tweets.

Contributing Authors, Gartner, Inc.

Mark Beyer, vice president and distinguished analyst, is the co-lead for Big Data research where he researches practical use cases in this area. He also covers traditional data warehousing, data integration and information management practices.

Frank Buytendijk, research vice president, covers "information innovation." Within this broad topic, Mr. Buytendijk specializes in information management strategies, big data and analytics.

Marcus Collins, research director, covers data architecture, data and information integration, database management system evaluation and selection, database architecture and emerging technologies (big data, NoSQL).

Jay Heiser, research vice president, specializes in the areas of IT risk assessment and management, collaboration security, security policy and security organization.

Anne Lapkin, former research vice president

Hung LeHong, research vice president and Gartner Fellow on the Executive Leadership and Innovation research team, focuses on senior executives and CIOs to help them anticipate changes to business models and consumer trends caused by technology disruptions.

Nick Heudecker , research director in Gartner Intelligence's Information Management group, is responsible for coverage of Big Data and NoSQL technologies.

Rita Sallam, research vice president, focuses on business intelligence (BI) and analytics.

Svetlana Sicular, research director, covers data governance, enterprise information management strategy and big data.

Chapter One: What Big Data Means for Business

By Doug Laney, Hung LeHong, Anne Lapkin

Big data is one of the most hyped terms on the market today. It's one of the most popular search terms by CIOs and other IT professionals on gartner.com, and according to the Global Language Monitor (GLM) it was the most confounding term of 2012. But while it may spell confusion for many, it also means big money for some.

As with every new term that creates excitement in the market, incumbent vendors are rebranding and expanding products, upstart vendors are saturating the landscape with big data solutions, and investors are swarming. True, many organizations *are* expending significant resources on big data projects; 42% had adopted big data technologies by the end of 2012 according to a Gartner study. But less than 15% currently have an enterprise strategy.

So what's the real enterprise promise of big data and how should CIOs plan strategy, execution and resources to improve their operations and seize competitive advantage for their organizations? What are the risks to CIOs and their organizations that ignore the wealth of data that is now at their fingertips? Gartner experts will answer these questions in this eBook single for the *Financial Times* Connected Business.

The Big Data Difference

Consider this: A telecommunications company wants to reduce its risk of customer loss, so it analyzes billions of call detail records to find out which customers are the most connected (that is, make or receive the most calls from a wide variety of phone numbers). The company then focuses promotions on these individuals to keep them as happy customers, since if they leave, they may "drag" a lot of friends with them to a new carrier. It is this type of hidden insight that demonstrates how big data expands the range of information used in decision making. Enterprises can now create new business value by leveraging sources of data that were previously hard to capture, access and analyze because of challenges with its size, speed and structure.

To get to the heart of what big data means, here's Gartner's definition:

Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.

The value in accessing the ever-expanding pool of data is great. However, it hasn't been easy for CIOs and their organizations to analyze it all. There's been too much of it; sometimes it comes in at real-time speeds, and the tools to analyze different types of data (for example, video and social feeds) did not exist. Now, new innovative and cost-effective technologies make it possible for organizations to handle these challenges, opening up a whole new realm of possibilities.

Three Categories of Business Opportunities

Big data can unlock new business value in a wide variety of ways, but most notably in three types of opportunities: Making better-informed decisions, discovering hidden insights and automating business processes.

Better-Informed Decisions

In the first case, decisions such as prices, promotions, staffing levels or investments — any business decision — can be improved if big data sources are available for insight. Take for example, Wal-Mart, which wanted to help its website shoppers find what they were looking for more quickly. It developed a machine learning semantic search capability using clickstream data from its 45 million monthly online shoppers combined with product- and category-related popularity scores generated from text mining social media streams. Wal-Mart's resultant "Polaris" search engine yielded a 10% to 15% increase in online shoppers completing a purchase (or around a billion dollars in incremental sales).

Hidden Insights

Big data analysis can also be used to discover opportunities that are obvious only by looking at large sets of detailed data. Many organizations are mining vast pools of data to discover hidden insights that were previously unavailable to them—often in the development of new or enhanced products.

The Climate Corp was started by former Google employees to offer crop insurance to underserved parts of the world. It continually gathers weather and soil measurements from 500,000 locations and has 30 trillion data points to date. Complex analytics predict weather-related risks for specific crops in specific locations. This enables Climate Corp to outcompete other insurers that cannot assess risk at that level of locale specificity, and enables farmers in Asia and Africa to take on the risk of buying seeds, labor and farm equipment that they otherwise could not.

Automate Business Processes

Finally, new technology can be used to leverage big data in real time, allowing analysis to be built into processes so that automated decision making can occur. One of McDonald's bakeries replaced calipers and color cards with high-speed image analytics to scrutinize thousands of buns per minute for color, size and even sesame seed distribution — instantly adjusting oven and other process controls to create uniform buns and reduce wastage. Another food products company similarly photo analyzes and sorts each and every French fry produced to optimize quality.

Tapping Into Dark Data

An enterprise that is proficient in analyzing big data has a whole new world of data sources available. It can now leverage data both inside and outside the enterprise that was previously unavailable or not utilized. For example, inside the enterprise, there exist underutilized datasets, or "dark data." These can include email archives, warranty forms, call center recordings and doctors' notes. Large public sources of data such as social media and government also become a source of potential value.

More Than Technology

CIOs and their organizations have to invest in more than technology to get value from big data. Innovation, cultural change, analytical mindset and new skills are required to be proficient at leveraging it for the enterprise. A key area to address is the elevation of information to the new status of corporate asset. Yes, as a fuel for business performance big data is like oil or coal.

By 2016, Gartner predicts 30% of businesses will be wielding their information assets also as a currency — bartering or trading with them, or even outright selling them.

The technology is not the big problem — big data technologies are available from many vendors — and they can be acquired on-premises and in the cloud. The big problem is skills. Gartner has predicted that by 2015, four million new jobs will be created, yet only a third of them will be filled. The shortfall will primarily be in the data scientists — those professionals with enough business knowledge to ask sensible questions — and the statistical and analytical skills to structure and analyze the data to derive sensible answers. It is likely that these skills will not all be resident in a single individual, but rather CIOs would do well to structure big data teams, looking in unconventional places to staff them. For example, sentiment analysis of social media streams requires linguists who understand how language is used.

Vision is also a significant challenge. In Gartner's 2012 Survey of CEOs and Business Leaders, fully 40% of the respondents had no idea what types of information would be disruptive in their industry in the coming decade. These executives may find it hard to imagine what questions to ask since they've never been able to ask them before due to technological or cost constraints. Working with business executives to help them envision the art of the possible is increasingly part of the CIO's job. But CIOs and business leaders who invest early and wisely in big data initiatives can give their organizations a competitive advantage and lock up partnerships and recruiting ahead of the market. We'll explore how in the following chapters.

Chapter Two: Build a Big Data Strategy Based on New Digital Connections

By Svetlana Sicular and Marcus Collins

While discovering hidden insights and making better-informed decisions means we can now predict fires in the Amazon rainforest six months before they occur by analyzing sea surface data, help arrest a shoplifter who tweets outside the store by analyzing social sentiment and geolocation data, and determine whether surgery is the best course of action for a patient by assessing trends across large numbers of lung cancer patients; generating ideas and knowing which ones to pursue is difficult. One way to get started with big data is to conceive scenarios of extracting insights for decision-making and operational efficiency by taking advantage of the "four Internets." These separately identifiable, virtual Internets of people, of things, of data and of ideas are emerging to enable broader collaboration and knowledge. They are also an invaluable source of big data to fuel advances in business capabilities.

The Internet of People

The Internet of People is represented by the set of interconnected information about individuals, including their social and collective activities and interests, their attitudes, and their images, audio and video. This can offer segmented and holistic views on human behavior, perceptions and interactions in space and time. At its core, it is about customer centricity.

Businesses can explore the use of big data by asking the ultimate question, *What can I do together with you?* Instead of the more traditional, *What can I do to you?*

The Internet of People allows organizations to expand business processes beyond the borders of the enterprise. This enables the fashion industry, for example, to find the next craze before it occurs by analyzing what people talk about in social media. Or, companies can invite customers to help solve problems and exploit opportunities by giving them rewards and incentives or find what makes individuals more positive or negative and adjust the business accordingly. In another scenario, businesses consider what they wanted to know all along about the customer if they had unlimited capabilities. Big data technologies can help find patterns for areas such as, who are the customers of our customers? What do our patients, accountable for the highest costs, have in common? How do we make connections between seemingly disparate people, places and events to detect fraud?

The Internet of Things

The Internet of Things is the data that represents the connections between the physical and digital worlds. It is growing at an unprecedented rate because of the lowering cost of the components that are turning "things" into parts of a network.

Whenever there is a possibility to get information about a physical object or a process by instrumenting it with sensors, RFID tags, transmitters, GPSs, logs and

other means of sending information via wired or wireless networks, there are opportunities to analyze the data and find new patterns.

Sensors can transmit information from the hardest-to-reach places, such as a working engine, a human body or a pipeline segment in a remote location. McKenney, the mechanical contractor firm, developed Business Intelligence for Buildings by tracking trends and performance over extended periods. It optimizes the uses of energy, water and indoor air quality across hundreds of buildings to achieve new levels of cost efficiency, such as reducing energy usage by 5% to 10%. In another example, UP Jawbone is a personal system that combines a wristband and phone application to track how people sleep, move and eat to know themselves better and make smarter choices to feel better.

To utilize event-driven data from things, companies explore how to prevent undesirable events such as device breakdowns, traffic jams or cyberattacks. They can also assess how to maximize positive events. These include areas such as reordering parts, administering medication or finding parking on a busy street.

The Internet of Data

The Internet of Data is about bridging information silos to understand physical, societal and business environments. It achieves this by connecting data at scale, both inside and outside the enterprise. The most obvious characteristic of the Internet of Data is variety: text, logs, images, video and geolocation, combined into a data fabric, to hold the information that organizations wanted to have all along. The accelerating liberation of data is the sign of a more open society and, consequently, more open and available information. Many governments provide data about demographics, economy, weather and the well-being of their citizens. Commercial entities seek to monetize their own data.

Companies should look for data-derived opportunities by detecting behavior in groups, fraud or life cycle patterns to gain new or even breakthrough insights. For example, because of linking and analyzing longitudinal patient data, family history, genetics and reference data, a healthcare provider can discover a new treatment for a particular patient based on treatment results for "similar" patients. Or, companies can develop data-driven business models and information products by combining their own data, data sources from partners, and purchased information or open data. It's also important to drive business strategy by making data-driven choices and find where evidence-based analysis can substitute or complement a "gut feel."

The Internet of Ideas

The Internet of Ideas is about the power of connected minds. It involves humans at scale and aggregates individual ideas about societal, business and physical environments through crowdsourcing, crowdfunding, leveraging open-source products and integrating ideas from outside the enterprise. In 2000, Goldcorp made an unprecedented decision to open up its proprietary geological data for the "Goldcorp challenge," a public competition to find gold in its Canadian mine. Out of the top five entries, four have been drilled, and all four struck gold. Since then, Goldcorp has grown from a $100 million company to a $9 billion company.

The complementary strengths of humans and technologies are mutually reinforcing. Opportunities for finding cost-effective solutions that involve human touch include business processes where instead of separating people, organizations combine human and machine

intelligence for better outcomes or make decisions by relying on machine analytics. Think of the two mediocre players who used a laptop running a commercial chess program to best the chess machine that had beaten a grand master on its own.

The Internet of Ideas also provides solutions for getting new ideas from outside sources or that need multiple perspectives or statistically significant representation of participants. The company, Factual, maintains a crowdsourced definitive database of 66 million local business and point-of-interest listings across 50 countries. In this case, submissions by thousands of people create detailed information that could not be obtained without individual inputs on a mass scale. Organizations can treat human minds as an equally possible analytical and business solution to find scenarios where they can benefit from crowdsourcing, crowdfunding and the expansion of enterprise borders.

New opportunities require the mental shift toward accepting big data realities. Organizations must revisit the problems that were once impossible or impractical to solve: The answers were contained in the data all along, but they were hard to extract with old technologies — it is doable now. When organizations allow themselves to ask bigger questions of people, things, data and ideas in today's interconnected world, they can find new answers to derive business value from big data.

Chapter Three: The Art of Big Data Innovation

By Frank Buytendijk, Research VP — Information Management

In business, we often deal with hype around trends in society, politics, economy and technology. We know we need to take claims of the next big thing with a grain of salt and that we should be careful not to set expectations too high. However, with *big data*, the opposite is true. The hype that accompanies it actually conceals the enormity of its impact on the way we do business.

In order to understand the value of big data, it is important to realize that it is neither merely about "big" nor about "data." The "big" part refers to the high-volume, high-velocity and high-variety nature of the information assets. But volume is only one aspect, and in many cases not the most difficult issue to overcome. The same goes for the velocity with which the data flows. The technologies to handle this may not be familiar, but they do exist. Instead, most of the value and challenge is in getting the most out of the new *variety* in the data. Current business cases in big data show meaningful combinations of smartphone location data, video feeds, internal process data, text documents and the weather forecast; just to name some of the various types of big data available today.

When it comes to the "data," it's important to realize that the emphasis is on a different understanding of the value of information.

This occurs by shifting from a traditional top-down sense, with an existing business question in mind that requires an answer, to valuing new ideas and new opportunities that emerge from all kinds of data, through a process of induction.

Where Big Data Succeeds

Gartner called 2013 the year of experimentation for big data — a year in which companies other than Internet retailers, large and small, start to discover the value of big data for their organizations. During this time, hundreds of useful business cases have emerged. Clearly, organizations see the biggest potential in improving customer insight and interaction. But when looking at where investments have gone so far, process improvement leads the way.

Operational Excellence

It makes sense to start exploring big data opportunities for process improvement before moving to customer value. It is always a good idea to have your shop in order before you start advertising.

Asset-intensive industries, such as telecom, manufacturers, utilities and transportation, can build a strong business case to outfit equipment with sensors that help with "predictive asset maintenance." By measuring vibration, sound or variations, maintenance on equipment doesn't have to come by surprise. The cost of unplanned maintenance can be driven down. This is an area that is called operational technology, or OT. The leaders in this area are not your typical IT companies, but they come from manufacturing. Think, for instance, of GE and Siemens.

Especially GE has made a dramatic strategic move, building a value proposition based on sensor-based data streams, instead of positioning the specifics of a new engine or windmill.

Other existing business cases receive new life from big data such as sales and production forecasting. More information about target audience demographics, the weather and social media activity on certain topics may improve the quality of the forecast, as well as increase the periodicity. Monthly forecasts may become daily ones, or even become event-driven.

Customer Intimacy

By far the most widely recognized business case for big data is sentiment analysis. Through structured analysis of unstructured data such as social media, organizations can determine their reputation or that of their products and services in the market, and get clues on how reputation may be changing. Sometimes this can go very quickly, and immediate response is required. Targeted advertising and recommendations are another popular category.

Sentiment analysis has become such an accessible analytic, and there are so many service providers offering it, that it is hard for any organization, public sector or commercial, to justify why they are not using it.

Risk Management

Big data is very often used for fraud management. Graph analytics, in particular, can help in detecting fraud rings. By understanding relationships in the data, hidden commonalities can be discovered. In some instances, people who commit fraud share a certain location, are from a similar age group, or are seemingly disconnected companies sharing common ownership. The results are real. Not only in terms of business value, but they also lead to societal impact. The newspapers recently reported that in Europe, where healthcare is often less privatized, various insurance companies have uncovered claims fraud by dentists and other healthcare providers who were using diagnostic codes that didn't match real procedures.

Banks often use big data for improving credit scoring, using graph analytics to include social relationships as an indicator of risk, or better yet, credibility. In an ironic twist, modern technology is reinventing old values in community banking to provide knowledge that a certain family is good for its money, even though family members growing up haven't shown that behavior themselves yet.

New Business

Perhaps the most exciting business cases come from a new value discipline: treating information as a product in itself. Utilities and banks already provide their customers with personalized dashboards about their use of financial products or energy. Remote patient monitoring is a growth business for healthcare providers or life science companies. Wearable computing introduced a category called "personal analytics," where consumers can measure and share health indicators such as heart rate, blood pressure and calorie consumption.

Industry Inspiration

Success falls to those organizations who creatively embrace big data. However, it's also important to note that inspiration often comes from other industries. A big data scenario from travel and transportation can, for instance, be used in retail for automated store task management.

In this instance, a train company uses video feeds of its cabin safety cameras to provide information to travelers at the next station about which cabins still have enough seats available. A retailer can leverage this to count how many people enter the supermarket, combine that data with an understanding of the average shopping time on a weekend in rainy weather conditions to help predict when to open up a new cash register.

In another example, malls can learn from the games industry. By tracking through a smartphone where mall visitors are, coupons can be shared on the spot. Additionally, visitors earn points by tracking how many stores they visit and are promoted to the next level.

In conclusion, big data is not about just handling volume, nor is it about data. It is about creativity. Combine technology advancements with human ingenuity and the possibilities are endless.

Chapter Four: Big Data Strategy Essentials for Business and IT

By Doug Laney and Mark Beyer

Big data initiatives are all about change — changing business processes, data sources, infrastructure, architecture, skills, organizational structures and economics. And they often result not in incremental improvements to existing business processes, but in radical changes to existing processes or even their outright displacement. Business leaders need to be involved in laying out the big data strategy for the enterprise but CIOs need to take the lead and help business leaders see how to use big data and ensure that the infrastructure and skills are in place to capitalize on it. Together, business and IT need to adopt a set of essential strategies as they embark on big data initiatives for the enterprise.

Recognize How Big Data Initiatives Are Unique

Business executives should consider that big data projects tend to concentrate on acquiring, integrating and preparing information rather than the data's functionality. This shift in focus can strain traditional approaches to enterprise architecture, project management and role definition. Another perceptible difference with big data projects, and the one we believe is given a disproportionate amount of press as a result, is the underlying technology. Traditional, even state-of-the-art, hardware, database management systems and analytics capabilities are often dispensed with in favor of technologies specific to accommodating massive, swift and diversified data and analysis. For those indoctrinated in the traditional ways of data warehousing and business intelligence (BI), these changes can be arduous.

Additionally, big data initiatives require a degree of financial rumination and discipline focused on the question, *"What value can we generate from this data, and is it more than it costs us to accumulate, administer and apply it?"*

The outcome of big data projects can be uncertain. Even more uncertain is the ability of many businesses to act on what they find in the data.

With time being money, how quickly can your organization get from focused experimentation that yields insights or innovations to its implementation and institutionalization?

Generate Big Ideas for Big Data

Business sponsors must also realize that major opportunities involve ways to transform the business and disrupt the industry by asking and answering "chewy" questions that were never possible before. We discussed how to generate big ideas in our earlier article, "Building a 'Big Data' Strategy." Building on that, business executives should ask questions that go beyond the mundane types of questions answered by basic BI tools such as, *"How much did our business grow in the past year?"* Instead, they should ask questions that make full use of broader, deeper and more real-time data and, if answered and acted upon, could have profound effects. For example: *"How can we increase customer shopping basket value by 20% and loyalty by 33% by better understanding their individual interests and behavior, and considering a range of economic forecasts and competitor moves."*

Build Business Leadership Belief in Data

Unfortunately, many business leaders are still resistant to relying on data for decision making. Especially in matters of strategy, deep personal or professional experience, or multidimensional factors; business leaders rely on intuition more often than benefits to their organizations. In strategic decision making, leaders tend to overemphasize past individual experiences despite new or differing data indicating situational change. Even more common today, as information becomes more complex and analytic techniques become more sophisticated, is the inclination merely to discount data or formulae that one doesn't understand.

Some of the remedies to this discounting of available data that CIOs should enact and business leaders should embrace include executive education in basic statistics, risk/scenario planning, "group think" avoidance and even decision theory; decision competitions among individuals or teams; communicating analytic insights and their transformative opportunity; or pairing data scientists directly with executive teams.

Embrace Investment Pragmatism

Big data doesn't dramatically alter the economics of acquiring, administering and applying information assets but it does amplify it. No longer can organizations ignore the need to balance these information supply chain costs with the tangible value derived from information.

One important scheme for tipping the balance of big data benefits to outweigh its cost is ensuring that the data serves multiple business purposes.

Compiling, hosting and processing petabytes of data for a single business process rarely makes for sound financial fundamentals or good use of scarce skill sets. Although many Big Data investments may start out as either speculative experiments or focused entrepreneurial efforts, ultimate strategies should include expanding the utility of the data, algorithms, skills and technologies for additional business functions.

Ensure Infrastructure Adequacy

For its part, IT needs to ensure that the technology infrastructure is sufficient for the multifaceted demands of big data. Many traditional and even state-of-the-art technologies were not designed for analytical processing or traditional data warehousing — at least not for today's or tomorrow's combinations of data volume, velocity and variety.

IT generally defaults to extending existing systems capabilities to meet new processing demands. But since generating business value from big data is so urgent and potentially impactful, merely waiting on technology evolution is sometimes not an option. The strategy to "extend" needs to be tempered by knowing the limits of one's current technology portfolio when larger, faster and more diverse data needs to be managed and analyzed. Investing in new purpose-built technologies may be necessary.

Prepare for Business Risks

Big data also raises the specter of significant risk to business brand and compliance. Data sources frequently include personal, sensitive or proprietary information that can be more prone to mishandling and misuse. Even when individual data sources themselves do not contain explicit information, the integration of multiple sources may enable triangulation, or a so-called "mosaic effect," that could expose corporate secrets or identify individuals. This risk can be especially perilous when information is intended to be shared outside the organization with

business partners, suppliers, customers, trade organizations or government. Therefore, Big Data strategies, as rogue as the efforts may be, must consider governance, controls, monitoring and even contingency plans.

Expand Existing Analytic Skill Sets

Analytics is the No. 1 use of big data, yet common BI solutions are limited in their analytic capacity — particularly with unstructured data or analytics beyond hindsight-oriented reporting and extrapolation. Big Data efforts demand looking beyond traditional query and reporting capabilities to consider predictive analytics, data, text and even multimedia mining, increasingly illustrative and layered forms of visualization, complex event processing, rule engines and natural language query.

Understanding how to apply these capabilities demands a range of skills, but the new talent required to manage and leverage these information assets is in exceptionally short supply.

These skills include data integration and preparation, business and analytic modeling, collaboration and communication, and creativity. The role of the data scientist is emerging as somewhat of a panacea, not only for generating new insights, but also for finding ways to use available data in automating and optimizing business processes.

Alter Organization Structures

Big data initiatives have a strong tendency to stretch and test traditional IT organizations in unique ways. Most are badly equipped to deal with an individual business unit's desires or attempts to manage and leverage big data on its own—often outside the context of traditional data warehouses and business intelligence efforts. CIOs must be prepared to affect the necessary changes because resisting them to maintain IT standards and the status quo will result in being shut out of enterprise strategy dialogs. Because big data initiatives are especially demanding on the partnerships between IT and the core business, it's essential that both groups weigh the necessary strategies and planning to maximize the organization's return on big data initiatives.

Chapter Five: Infonomics: The New Economics of Information

By Doug Laney

Most businesses are frantically curating and leveraging information to improve business performance and innovation. But while the race is on to innovate with big data, a large chunk of a company's information assets are unreportable on corporate balance sheets. In other words, despite information arguably meeting the formal criteria of what constitutes an asset, the keepers of the "asset torch" maintain the antiquated notion that information is not an asset. This incongruence has confounded legal systems worldwide (is data property or not?), and hampers the intensifying enterprise imperative to manage information with the same discipline as acknowledged assets.

The problem stems from archaic and arcane accounting practices that disallow the capitalization of information assets and make big data's large and fast-growing swath of corporate value unaccounted for on an enterprise's books.

And since information is not accounted for as an asset, the insurance industry refuses to recognize it as property. In fact, after 9/11 — when some companies attempted to submit claims for the value of the data they lost — the U.S. insurance standards body, Insurance Service Office (ISO), revised the Commercial General Liability Policy template to explicitly exclude coverage for information assets. When did they do this? One month after 9/11. Not to be outdone, the accounting profession followed suit a couple of years later by revising FAS/IAS 38 to explicitly restrict recognizing most forms of corporate information.

Does Big Data Equal Big Value?

Today, most organizations are challenged by dealing with data that is bigger, faster and more diverse than in 2001. But do these increased levels of data volume, velocity and variety necessarily correspond to an increase in value? Actually, it's not so simple. The answer depends on how you define value.

An issue often debated is whether information's value depends on its use, or whether, like any asset, it possesses an inherent value, used or not. We call this the information value gap. The truth is, information assets have both potential value *and* realized value. Acknowledging, measuring and closing this gap is crucial to the successful use of big data in any business. In addition, accounting standards define assets as having "probable future economic benefit" — which can be useful in determining a realistic rather than theoretical potential value of an information asset based on the organization's anticipated capabilities. This accounting definition of value is germane for companies to create supplemental, internal balance sheets for their information assets.

So the answer is no, big data does not inherently equate to big realized value, but yes, it does equate to big probable/potential value. Mind the gap!

Measuring the Big Data Value Gap

If accountants and insurers don't care about the value of information assets, why should executives like you care? Well, big data begets big investments. Curating, managing and leveraging higher volumes, velocities and varieties of data, such as image, video and machine log data, demands new forms of information processing well beyond the traditional data warehouse and business intelligence environment. This includes storage, databases, computing power, analytics software and communication bandwidth, along with premium skills to make it all work.

Also, a lot of big data emanates from external sources, often those that must be licensed or purchased from data syndicators/aggregators, including credit data, consumer profiles and social media feeds.

Like any outlays, these investments in big data usually need to be justified regardless how speculative they may be. This is when methods for quantifying the cost and value of data come in quite handy.

One approach to value information is to borrow asset valuation methods from the accounting profession; namely, the cost approach, the market approach and the income approach. This entails determining the cost to acquire (or reacquire if lost) an information asset, its price in a real or presumed open marketplace, or its contribution to a revenue stream. These methods are fairly straightforward — certainly compared to other kinds of recognized intangibles, such as copyrights and patents. However, some organizations aren't quite ready to take the leap into quantifying the economic value of their information assets. They merely want to establish information-related IT or business priorities. In these cases, it's helpful to use valuation models that consider an information asset's comparative intrinsic value, its relative business value or its empirical performance value. These are based on key data quality metrics scope of business relevance and/or observed impact on non-financial indicators.

Closing the Big Data Value Gap

Measuring the potential versus realized value of data is just the first step — albeit a big step — that few companies have taken. Closing this gap demands that information is treated with the same discipline as other established assets, e.g., financial, material and certain intangibles.

In addition to solid, established information management and data governance practices, information managers, architects and strategists still have a great deal to learn from the way their colleagues in other departments manage traditional corporate assets. Concepts such as inventory management, planned/unplanned maintenance, supply chain management, portfolio management, and even organizational approaches all offer more than just morsels of principles and practices in asset management that can and should be applied to information asset management.

Introducing Infonomics

Infonomics is a concept that brings all this together. It is the emerging economic theory of information as a new asset class, and the discipline of accounting for, managing and deploying information just as any other enterprise asset. These notions are important in the context of any and all data, but are even more crucial for big data.

Key principles of infonomics include:

- Information is an actual asset
- Information has both potential and realized value
- Information's value can be quantified
- Information should be internally accounted for
- Information's net realized value should be maximized
- Information's value should be used to help prioritize and budget IT and business initiatives
- Information should be managed as an asset

Ultimately, executives who just continue to talk about information as one of their company's most critical assets, yet continue to eschew measuring and managing it as one, are doomed to continue having underperforming information assets. That's a big risk when it comes to big data innovation. It may mean underperforming businesses as well.

Chapter Six: Confronting the Privacy and Ethical Risks of Big Data

By Frank Buytendijk and Jay Heiser

One result of NSA whistleblower Edward Snowden has been the global surge in discussion about privacy and big data. This dramatic news story has contributed to increasing awareness over the use of big data by commercial enterprises to target and profile customers. The concern over whether governments are illegally collecting big data about their citizens reminds both organizations and individuals to consider the delicate balance between the benefits that big data analytics bring, and the ethical and privacy risks they pose. Individuals are not without responsibility by offering their personal data for free Internet services. Yet organizations should initiate an internal debate on the limitations of big data analytics and guidelines to avoid public embarrassment, mistrust and liability.

Big data, like most innovations, is a double-edged sword. It brings huge benefits. It allows organizations to personalize their products and service on a massive scale; it fuels new services and even business models, and can help mitigate business risks. At the same time, allowing data scientists to run amok can harm individuals and institutions in unanticipated ways. Notably, we predict that through 2016, 25% of organizations using consumer data will face reputational damage due to inadequate understanding of information trust issues, and 20% of CIOs in regulated industries will lose their jobs for failing to implement the discipline of information governance successfully.

Real Concerns

There is a subtle balance between improvements in operational risk and strategic risk by using big data techniques and increased reputational risk if (inadvertently) overstepping certain legal or social boundaries. There is an equally subtle balance between improvements in customer service and business operations by, for example, accurate customer profiling based on a variety of data sources, including social media and mobile phone data, and knowing so much that customers experience a "creep factor."

For example, if a pharmaceutical company analyzes DNA data, lifestyle data and socio-demographic data, on the lowest level of granularity, and draws interesting conclusions about health perspective then shares this data. But not sharing those insights could be unethical as well. What should the pharmaceutical company do?

Being responsible with big data is broader than addressing privacy concerns.

In an emerging field with so many possibilities, and where technology limits have shifted so dramatically, the consequences of use cannot always be foreseen. More than guidance and rules, a debate about what is principally right and wrong is needed. The following risks highlight what organizations must consider to protect their constituents, and themselves, from the impact of big data.

Risk 1: Anonymization and Data Masking Could be Impossible

Large datasets are often subjected to an "anonymization" process to enable the data to be used for marketing or scientific research, without the potential of leaking information about the individuals. However, no useful database can ever be perfectly anonymous.

Furthermore, for several decades, the information security research community has recognized that bodies of low sensitivity data, when they can be correlated, can often result in a set of data that has much higher significance than any of the original datasets.

When done with malicious intent, this is referred to as an inference attack, or slightly the more neutral term "reidentification." The "triple identifier" of birthday, gender and zip code is all that someone needs to uniquely identify at least 87% of U.S. citizens in publicly available databases.

The individuals who might have given permission to have their data used in what they believe to have been an anonymous fashion might have no idea that reidentification is even possible. This can lead to harmful results, revealing information on medical history, personal habits, financial situation and family relations that most people would classify as private.

Risk 2: Protecting People From Themselves

Not everyone cares enough about their own privacy. Many consumers use social media or Internet-based services carelessly, allowing others to make use of information in unintended ways. Consider the following examples:

- Publicizing on Twitter that you are on vacation or "checked in" somewhere with the whole family shows you are not at home.
- Consumers almost never read the "terms and conditions."
- To receive a promotion, consumers often need to provide some personal information.

Even though people are expected to know what they are doing, and there may be no legal issues after consumers consent to providing information, there is reputational risk to companies if consumers feel their trust and confidence was breached. What consumers trust you to do (or not do) doesn't necessarily equal what is legally allowed to do.

Risk 3: It's Easy to Mistake Patterns for Reality

Mass shootings, for example, in the U.S., have generated interest in attempting to determine which individuals are likely to act out on violent impulses. These clues are believed to be available in Facebook and other social media. Institutionalizing this type of activity could result in a sort of "Minority Report" phenomenon.

Governments are already conducting data mining of cash transactions to infer the activities of terrorists and other organized criminals. Police forces use advanced predictive analytics to predict a higher chance of crime rates in certain areas on certain days or times in the day. Surveillance cameras in streets are connected to analytical software that is engineered to detect behavioral patterns indicating trouble.

This may easily lead to "fishing expeditions," where authorities conduct mass analytic exercises, in which any person fitting a certain pattern becomes a suspect. For crime prevention purposes, there is a direct issue with the constitutional presumption of innocence. In business, a pattern doesn't necessarily equate to behavior.

Risk 4: The Data Becomes Reality Itself

In business, unintended behavioral influence happens as well. Based on advanced analytics, retailers provide customers with personalized offers. Confronted with perceived endless choices in online and street retail and a lack of ability to compare with other offerings, customers are likely to welcome such offers. The acceptance of the offer refines the profile, leading to an even more targeted offer, leading to higher conversion rates again. Through this closed loop, the profiling and associated prescriptive analytics start driving customer behavior, rather than the other way around. This is commercially interesting, but ethically debatable.

Risk 5: Don't Worry About Bad People; Worry About the Ignorant Ones

Big data analytics distinguishes itself through the use of automated discovery techniques, presenting potentially interesting clusters and combinations in data. This is a powerful tool when dealing with high volume and high velocity data with a high degree of variation, but also potentially dangerous. Customer segmentation and profiling can easily lead to discrimination based on age, gender, ethnic background, health condition, social background, and so on. These are limitations known to analysts, but not to technology. Knowledge, once gained, cannot be undone. Even deciding not to do anything with the knowledge is a decision with consequences already.

To guide big data analytics, it makes sense to also consider what data and analytics you would like to have, and equally important, what not.

One bank, for instance, removed face recognition algorithms from its set of analytics, because it didn't even want to be seen as being able to use it.

Organizations need to evaluate the value of knowing the answers to specific information-driven questions, analysis and models before they develop the model. Intent becomes the precursor to big data analytics. "Why do you want to know it" becomes the gateway before "what do you want to know."

Initiate Debate by Posing Ethical Dilemmas

To assess these risks, organizations should embark on an ethical debate of the arguments for and against certain actions with big data. By analyzing either their own initiatives or real world case studies such as the Snowden affair or Google's StreetView collection of data from private wireless networks, they can analyze multiple points of view on what is right and wrong.

Here, context is really important. What is acceptable for one organization might be unacceptable in another context. For example: Is it acceptable that a municipality, responsible for administering unemployment benefits analyzes social media data to check for fraud? Most people would agree to that, even without the consent of the people receiving the benefit. Should an employer set up big data analytics to monitor social media on staff behavior? More people would object, and staff would at least have to give consent. Is it a good idea for an insurance company to profile customers based on their social media data, analyzing communications on

their sports activities, dieting and smoking habits, and using the information for individualized premiums? Many people would object.

Ethical debate is about determining what is "appropriate" and what is "not appropriate" for your organization and for others. This is normative and subjective in nature. People have different values, principles, beliefs and convictions. There are differences between cultures and age groups. Fundamentally, an ethical debate forces an organization to take a stand, and determine what it believes itself to be — good and bad — instead of relying on regulatory compliance and industry best practices. The outcome should be implemented and repeatedly communicated and enforced.

Develop a Code of Conduct

As a result of this ethical debate, information leaders should, together with their marketing and legal departments, develop a code of conduct for big data analytics. This code of conduct should contain the list of principles that describe what the company finds appropriate and inappropriate, a process that describes the ethical checks and balances when conducting big data analytics, legal implications, whether the intended use of the data matches how it is actually being used, and if the organization would be comfortable if the results of it became public.

Experimentation has its risks. The more the analytical use of data is removed from the original goal of measurement, the higher the chance data is used in a questionable way. Use data in the way that its original measurement was intended for. Invest in metadata that describes the origin of data, the purpose of data and limitations for its use. And keep communicating the code of conduct. Ethical guidelines require regular attention and reinforcement. Make sure they are part of every business case and proposal and are part of a checklist when running campaigns or other analytical activities, send out reminders to critical staff, and solicit feedback from people involved on how the principles have been helpful. Guidelines are only effective when they are top of mind and enforced.

Chapter Seven: Key Trends in Big Data Technologies

By Rita Sallam, Mark Beyer, Nick Heudecker

Not long ago organizations convened focus groups to assess customer interest. Today, executives can ask their data specialists to find those insights in social sentiment, sales numbers, web site behavior, sensor data, and more. In the next few years, the ability to find and assess trends, turn insight into foresight, to tap the behavior of multiple audiences, and to optimize decisions will move into the hands of the executives, their teams, and even individuals. In the meantime, organizations must evaluate big data technologies with an eye on what's possible and practical for seizing opportunities in the future and today.

Investment in analytics and the information management infrastructure to support it has been a top CIO investment priority for the past six years. Yet according to a recent survey, only eight percent of respondents say they have deployed a "big data" project to production.

Many big data deployments are still in the knowledge gathering, strategy and piloting phases.

And the overwhelming majority of big data initiatives are processing traditional data sources, like transactions or log data rather than a variety of data sources, such as social, email, voice, machine or sensor data. As organizations prepare for big data initiatives, they should consider the following key trends of investment driving growth.

Analytics Will Be More Pervasive

It's not just about big data – it's about ubiquitous data. Analytics and insights from analytics will move out of the hands of a select few specialists to be more pervasively accessible to non-traditional business intelligence (BI) users, customers, and even for personal use. Most current tools require users to know the data they want to analyze, know the questions they want to ask in advance and also possess specialized skills to initiate queries, build analytics and mathematical models and build visualizations using the tools. These skill sets are beyond those possessed by most business users.

But this is changing as business users increasingly demand consumer-like capabilities that allow them to easily find causal relationships in data and allow them to use that as a basis to more precisely predict outcomes and prescribe the best action or decision to take (often in real time) to drive the greatest business value without specialized skills. Examples of scenarios on the horizon include routing a caller to the best call center agent based on the caller's voice sentiment, interaction history, social behavior and influence and demographics. To achieve the best outcome, the call center agent is automatically sent an optimized script, offers and treatment recommendations for that specific caller. Or, carpets equipped with sensors that monitor and analyze senior citizens' activities for dangerous abnormalities, which are then delivered via mobile devices with prescribed intervention or remedial actions for healthcare professionals and /or care givers.

These types of scenarios will become more mainstream over the next two to three years with technologies that give business users human friendly and intuitive visual interaction (for example, users would be able to initiate queries and analysis using natural language voice or text questions as inputs instead of having to access BI tools) and data exploration and discovery tools with guided recommendations for finding patterns in data and for conducting more advanced types of analysis. This will be achieved by embedding and encapsulating complex analytics from users, surfacing recommendations for optimal courses of action at the point of decision (increasingly on a mobile device), and incorporating the user's context (i.e. location, intention, sentiment, past behavior and network).

In addition, social and collaboration capabilities integrated with analytics will be increasingly important investment areas making it easier to share, discuss and socialize results and to provide a mechanism for making transparent, high quality decisions. Much like Amazon users are presented with a "people who bought this item, also bought this one" recommendation, analytics users will be presented with similar guided analysis based on the social profiles and decision history of other decision makers and their previous interactions.

Analytics Will Be More Precise.

Organizations are increasingly investing in capabilities that enable them to discover more precise patterns and micro predictions based on diverse data - increasingly in real time. This will require investments in advanced analytics for more precisely predicting likely outcomes with high productivity (iterating and refining many more models in a short period of time) and accuracy (on larger number of data dimensions) and in finding unknown patterns and relationships across the enterprise and within new types of data such as social, emails, call center interactions, video, and machine data. Examples include identifying fraud and cyber security threats, best next offer, predictive maintenance, predictive policing, personal monitoring for alerting and optimized healthcare, early identification of adverse new drug effects, etc. This requires new types of analysis such as sentiment, geospatial, and network analysis to find entities of interest, their relationship and influence. Organizations will also require new skill sets and may fill this gap by investing in a combination of internal skills building, outsourcing to analytics service providers or to crowd source analytic models.

Analytics Will Enable Better Decision Making

Decisions are a basic unit of work for all organizations. The success of every enterprise is a function of the cumulative effect of the quality of the decisions that it makes. Despite large BI investments in the name of better decision making, poor decisions are abound.

Where decision rules and logic are well known, more precise and real time analytics will be applied to automate a range of operational decisions.

For example, a retail food chain monitors refrigeration assets in real time to proactively predict and maintain an asset before it fails. At the same time, the quality of collaborative decisions and professional experiential and judgment-based decisions (clinical diagnosis, employee hiring, online education, personal health and wellness) will be enhanced by advanced analytics, man-machine partnerships or digital assistant models (think IBM Watson); and many more are emerging.

Moving toward something that looks simple and invisible from the user's perspective will require new types of computing capacity and power, extended capabilities and skills, and extended capabilities in information management systems, including but not limited to:

- Visual based data discovery
- Natural-language query so non-traditional analytics users can find insights in data.
- Contextual engines to understand the user context (for example, who users are, where users are, what users are doing, with whom are users interacting).
- Semantic technologies, text, speech and video analytics to derive new insights from previously in accessible data along with algorithms that simulate the way the brain understands, aggregates and relates diverse pieces of data, reasons, and learns — much like the human brain.
- Advanced analytics, such as predictive modeling, machine learning, graph analytics, sentiment analysis, statistics, and simulation and optimization techniques — including linear and nonlinear programming.
- In-memory computing, Hadoop, NoSQL, search technologies and event processing to handle large volumes of diverse and real time data.

Information Management Evolves

Analytics investments over the next three years will require an evolution of the information management architecture. Key enabling investment areas include data management hybrids for semantics and data integration, metadata management for transparency of source data, technology that relates business process model changes to the associated information assets, and graph analytics. Graph analysis is the best option for presenting alternative scenarios, scoring them and comparing them when combining highly disparate information asset types. It is not only possible, but highly likely that multiple connection points will exist between information that was not designed to be used together. Innovation requires comparing how to combine the same set of disparate information under multiple models.

Deriving value from big data likely involves several sources of data with varying levels of structure and relationships. Different analytical outcomes can be realized at different points in the information lifecycle. For example, continuously generating computations as data flows in, can yield insights in real-time; while analyzing the data in batches results in different outcomes. No single technology supports both types of scenarios. Therefore, big data technology choices will be driven in part by the physical and logical attributes of data in combination with the desired analytical or business outcome. Several different technologies must be combined when multiple outcomes are desired, such as real-time data processing and interactive data exploration. In the end, understanding what’s possible with big data through the technologies on the horizon will help organizations plot their course to innovation.

Conclusion

It may be convenient or even fashionable today to dismiss the topic of Big Data as nothing more than marketing hype with the purpose of selling IT organizations ever-larger and more sophisticated data management and analytic technologies. Indeed since the dawn of computing we have always had difficulty keeping up with the burgeoning availability and desire for data. However, reality suggests that the generation of information an order of magnitude larger, faster and more assorted than just a few years ago has become a principle economic driver. Evidence abounds that *infocentric* organizations – those managing and leveraging information as a real corporate asset—outperform and out-innovate their peers. Thirty years ago leading businesses were those that best took advantage of available physical assets; today's leading businesses are those that best take advantage of available information assets.

As my Gartner colleagues and I have advocated here and throughout our published research, this economic shift demands new analytic and technical skills, new forms of technology, and new types of managerial leadership. Surely an ongoing need for traditional data management and analytic technologies and skills such as data warehousing and business intelligence will persist. But the opportunities and challenges of Big Data extend well beyond these.

Even if the usage of the term "big data" diminishes, the growth and consumption of data will not. Look around you right now. Consider all the systems, devices, processes, objects or individuals within sight (and those that are not). They may not be spouting or devouring data today, but expect that they will be soon. Will yours be the business making this happen? Or will you be watching it happen?

--Doug Laney, VP Research, Gartner (@doug_laney)

For more information about Gartner, visit www.gartner.com