The Big Data Revolution: Should The Internet Learn To Forget?

Robert B. Milligan
D. Joshua Salinas

Biographical Information

Robert Milligan

Mr. Milligan is a Los Angeles, California based partner in the Litigation and Labor & Employment Departments of Seyfarth Shaw LLP. His practice encompasses a wide variety of commercial litigation and employment matters, including general business disputes, unfair competition, trade secret misappropriation and other intellectual property theft. Mr. Milligan has represented clients in state and federal courts in complex commercial and employment litigation. Mr. Milligan also provides advice to clients concerning a variety of business and employment matters, including non-disclosure, non-competition, and invention assignment agreements, cloud computing, social media, and privacy issues, corporate espionage, privacy, and trade secret and intellectual property audits.

Mr. Milligan is a member of the State Bar of California Intellectual Property Law Section Executive Committee and Chair of the Trade Secrets Committee. He is an ITechLaw member as well as the Chair of the Intellectual Property Committee. He is also the Vice Chair of the ABA Trade Secrets and Interference with Contracts Committee. He frequently lectures and writes on timely employment and intellectual property issues. He is the editor of his firm's trade secret blog, www.tradesecretslaw.com.

D. Joshua Salinas

Mr. Salinas is an attorney in the Los Angeles office of Seyfarth Shaw LLP, practicing in the areas of trade secrets, privacy, restrictive covenants, computer fraud, and commercial litigation. His experience includes the prosecution and defense of trade secret misappropriation, unfair competition, and consumer privacy claims. He is also an author on his firm's trade secret blog, www.tradesecretslaw.com.

Introduction

The Big Data revolution has created a new age of discovery and opportunity by equipping companies with tools to derive valuable insight from once unnavigable oceans of data. The rapidly accumulating and torrential amount of data produced every second from social media, mobile devices, and everyday transactions have fueled the rise of this revolution. The ability to analyze this wealth of information has allowed companies in a wide range of industries to make better real-time decisions and more accurate predictions in their daily business dealings. With those well-recognized benefits and the collection of personal and consumer information, there are incumbent risks to privacy. The current notice and consent privacy framework, however, is ill-suited for Big Data, which derives the most value from the secondary use of data that was initially unimaginable. Additionally, the ability to record such information for an indefinite period has raised the issue of whether people have an unbounded "right to be forgotten" on the Internet.

Companies employing Big Data strategies must ensure that their collection and use of data complies with the morass of applicable laws as well as carefully scrutinize agreements sharing "their" data or using "others" data. Additionally, companies utilizing third parties for Big Data services must employ effective measures to protect and secure their intellectual property. Vendor agreements with Big Data service providers and licensees should be carefully scrutinized to ensure that appropriate contractual provisions are in place to protect company data, including provisions addressing ownership, access, protection, and privacy

from both a national and international perspective. Sensible executives will seek advice from competent counsel to ensure that the invaluable opportunities in Big Data are not outweighed by the potential legal and business risks.

Robert B. Milligan
D. Joshua Salinas
Los Angeles, California, May 1, 2013

Big Data is the hot new phenomenon that is transforming the world in ways we have yet to understand. For example, some companies are using Big Data to identify and address issues within their organizations to lower their business costs, such as insurance premiums. For example, companies may analyze their internal complaints/incidents to discover whether the frequency or nature of certain complaints/incidents is correlated with a certain department, geographic region, employee position, major event, or other relevant data source. The companies then focus their resources and efforts to address these areas and reduce the likelihood of future issues through pro-active risk management, including employee training and technology solutions. Accordingly, the companies cite these pro-active activities to support their requests for lower insurance premiums or other reductions in cost.

In the retail industry, companies are using Big Data to personalize their marketing activities and better the customer experience. For example, a restaurant may discover that certain menu items are more popular in different cities or during certain times of the week. By understanding customer preferences and product demands, the company can make improve the individual customer experience and supply chain decisions that may improve market share and increase profits.

Social media, mobile devices, and everyday business transactions that produce massive amounts of data about our

everyday lives at an unprecedented rate fuel the ascent of the Big Data megatrend. Additionally, cloud computing has reduced the costs and facilitated the ability to store such large amounts of data. Big Data allows companies to navigate this ocean of data and mine valuable insight about how people live, work, and interact with each other. The main goal is to go from data to information to insight that is actionable or provides a higher level of understanding. This new phenomenon has already shifted certain traditional ways of thinking in many industries, including retail, government, education, finance, insurance, and healthcare. Indeed, Big Data is creating a new age of discovery into human behavior that was previously unimaginable.

Accompanying the emergence of Big Data is the rise in open source tools and new commercial tools in the data mining market. According to a new market research report, the total Hadoop[1] market is expected to reach $13.95 Billion by 2017 with a CAGR of 54.9%.[2] The Hadoop market in 2012 was worth $1.5 billion and is expected to grow to about $13.9 billion by 2017, at a CAGR of 54.9% from 2012 to 2017. Other tools and methods used in Big Data include SOM (self-organizing maps), cluster analysis, classification trees, GIS (geographic information systems), text mining, and sentiment analysis. Recent research from the CMO Council also reveals that both CIOs and CMOs believe big data is

[1] A major open source software framework used to analyze Big Data.
[2] Hadoop & Big Data Analytics Market [Hardware, Software, Services, Hadoop-as-a-Service] - Trends, Geographical Analysis & Worldwide Market Forecasts (2012 – 2017), http://www.marketsandmarkets.com/Market-Reports/hadoop-market-766.html

a key competitive differentiator and essential to affecting a more customer-centric business culture.[3]

While Big Data is a popular new buzzword, many companies are still struggling to comprehend what Big Data actually is and how they can effectively utilize it. Big Data lacks a rigorous concrete definition, but is often described as "***an attempt to manage increasingly large and complex amounts of data that cannot be easily stored, processed and analyzed using traditional analytical tools.***"[4]

A common misconception is that Big Data is simply defined as a vast amount of data. Size, however, is not the only factor. Big Data is often characterized by the 3 Vs: Volume, Velocity, and Variety.[5]

Volume is the sheer amount of data produced or modified that must be collected, analyzed, and managed. "This is perhaps the most immediate challenge of big data, as it requires scalable storage and support for complex, distributed queries across multiple data sources."[6] For example, retailer loyalty card data was generally not processed and a significant amount of potentially useful information becomes ignored data exhaust. Videos from surgeries were often deleted within weeks to save storage space on local computers. Now each set of data is processed, analyzed, and

---

[3] http://www.cmocouncil.org/download-center.php?id=259#

[4] http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

[5] http://blogs.gartner.com/doug-laney/deja-vvvue-others-claiming-gartners-volume-velocity-variety-construct-for-big-data/

[6] http://www.cognizant.com/InsightsWhitepapers/Big-Data-is-the-Future-of-Healthcare.pdf

stored for future use. Cloud computing has also helped fuel the rise in Big Data because it reduces the costs of collecting and storing vast amounts of data.

Velocity is the rate data is being generated or modified, and the speed at which it must be collected, analyzed, and understood. For example, many energy organizations have implemented smart meters and to monitor high-volume events and make important real-time decisions to avoid blackouts or widespread distress on their electric grids. Manufacturers are also capturing events from their systems, applications, and other devices to analyze real-time metrics and take corrective actions before any failures occur. Companies that can analyze their data in real-time and make real-time decisions or take immediate actions often hold the greatest competitive advantage.

Variety describes the unstructured nature of data that does not fit easily into traditional analytics or spreadsheets. Structured data is highly organized and manageable. The most common form is a database where specific information is stored neatly in columns and rows (e.g. Microsoft Excel). Unstructured data, however, has no identifiable structure Examples include data generated from images and videos, GPS locations and movements, voice data, emails, blog posts, web sites, and emotion-laden keywords in social media posts. Indeed, the penetration of mobile devices in our daily lives has fostered a wealth of data reflecting how we live.

The sea of data made daily through social media is a great demonstration of these three Vs.[7] For example, Facebook has 1

[7] http://expandedramblings.com/index.php/facebook-stats/

billion monthly active users who **each day** generate 2.7 billion page "likes" and upload 350 million photos. YouTube has 1 billion monthly active users who upload **one hour of video every second**.[8] Twitter has 200 million monthly active users sending about 400 million tweets per day with 60% accessing via a mobile device. LinkedIn has 200 million users and receives two new members every second. Indeed, people are producing data at an alarming rate revealing an opportunity for high quality insight into their consumer preferences, interpersonal behaviors, and everyday personal activities.

With Big Data, the more data that is available, the better the decisions become and the more actionable the insight becomes. The underlying principle is to look at all the information available to discover the different relationships and connections that could not previously be seen. As IBM's Big Data expert Jeff Jonas says, you need to let the data "speak to you."

Big Data expands our traditional tools of data analysis. For example, the U.S. Census Bureau conducts approximately two hundred economic and demographic surveys a year based on sampling.[9] Sampling can be effective, but its accuracy is based on the randomness of the sample and broad assumptions about the population being sampled. Sampling loses its usefulness, however, when you want to take a closer look into detailed subcategories or relationships within the data.

Moreover, there are several risks associated with sampling. Sampling results always are accompanied with a host of

[8] http://www.onehourpersecond.com
[9] http://www.census.gov/aboutus/surveys.html

assumptions, caveats, and a confidence interval that usually limit the actionable effect of sampling. Systematic biases in data collection may lead to extrapolated results lacking accuracy. For example, during the 2008 presidential campaign between Barack Obama and John McCain, the major polling organizations Gallup, Pew, and ABC/Washington Post found differences of between one and three percentage points when they polled with and without adjusting for cellphone users.[10] The problem with polling only land-line phones tended to exclude data from the younger and more liberal people who only used cellphones.

Sampling is usually the method to use when you don't have all of the data or obtaining the data is too expensive or impractical. So, sampling does have its usefulness, but this usefulness is being eroded with the increase in data collection.

In his fight against cancer, Apple's Steve Jobs had his entire DNA sequenced, not just a sample of his biomarkers.[11] The goal was for Job's team of doctors to select and tailor therapies based on his specific genetic makeup. Whenever one treatment lost its effectiveness, the doctors could switch to another drug. This personalized and tailored approach is believed to have extended his life by several years.

Additionally, Google's Flu Trends[12] website can effectively monitor and track in real-time the outbreak of the flu by analyzing Google's billions of Internet search queries (not just a sample) for

[10] Big Data: A Revolution That Will Transform How We Live, Work, and Think, Viktor Mayer-Schönberger and Kenneth Cukier (2013), p. 24.
[11] http://www.time.com/time/magazine/article/0,9171,2097955,00.html
[12] http://www.google.org/flutrends/

flu related keywords. Google has shown to predict flu outbreaks faster, more accurately, and with more location details than the Center for Disease Control.

Big Data has numerous benefits for almost every other industry. In particular, Big Data provides insight into linked behaviors and activities and allows companies to make predictions with this newly discovered information. Big Data can tell you *what* is happening, but it does not give the *why*. Accordingly, "society will need to shed some of its obsession of causality in exchange for simple correlations…."[13]

The retail industry has been at the forefront of utilizing Big Data analytics. For example, Wal-Mart analyzed its past transactions to see what items each customer bought, what else they purchased, the time of day, and even the weather. Wal-Mart discovered that prior to a hurricane, Pop-Tart sales increased. So as storms approached, Wal-Mart stocked Pop-Tarts at the front of stores next to hurricane supplies, and saw a boost in sales.[14] The retail and marketing industries have also effectively utilized lift curves and ROC curves as specific measurements of retail success with data mining.

Target Stores was also able to detect when a woman was pregnant by analyzing her purchasing patterns. The key purchases were not pregnancy tests or certain baby items, but large amounts of unscented lotion at around the third month of pregnancy and

---

[13] Big Data: A Revolution That Will Transform How We Live, Work, and Think, Viktor Mayer-Schönberger and Kenneth Cukier (2013).

[14] http://www.nytimes.com/2004/11/14/business/yourmoney/14wal.html?_r=0

vitamin supplements. Target then uses this information to send relevant coupons or targeted marketing materials to these customers. In fact, a father in Minneapolis last year stormed into Target angry that the store was encouraging his high-school teen daughter to get pregnant because Target was sending her baby related coupons.[15] The man later apologized after he spoke with his daughter and discovered that she was in fact pregnant.

Aside from the retail industry, Big Data has led to new discoveries in health and medicine. For example, Internet search queries are being analyzed to discover the side effects of drugs and interactions before the FDA could even report them. In fact, scientists analyzing Internet search queries discovered that when paroxetine, an antidepressant, is taken with pravastatin, a cholesterol-lowering drug, the interaction raised the person's blood sugars.[16]

New York City had a problem where its 300 pound cast-iron manholes would randomly explode into the air several stories high before crashing down to the ground.[17] A data analytics team started with 106 predictors, and subsequently discovered that the age of cables near the manhole and a record of past problems for that manhole could predict the likelihood of the manhole exploding. Predictive analytics methods were especially helpful with this situation considering the city of Manhattan alone had approximately 51,000 manholes.

---

[15] http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/
[16] http://jamia.bmj.com/content/early/2012/12/24/amiajnl-2012-001234.extract
[17] http://www.wired.com/wiredscience/2010/07/manhole-explosions/

Other industries are capitalizing on the use of Big Data to make predictions from correlations. Shipping company UPS has been able to monitor its fleet of 60,000 trucks and predict when preventive maintenance is needed. Sensors are often fixed to bridges and buildings to detect signs of wear and tear. Law enforcement and Homeland Security are collecting and analyzing data from social media to detect national security risks. Financial credit scores are being used to predict how likely people are to take their medication. Some insurance companies have used credit reports and consumer-marketing data instead of blood and urine samples for certain applicants. The method uses lifestyle data including hobbies, websites visited, and amount of television watched to analyze the applicant's health risks. Some insurance companies are also providing free plug in devices to analyze driving habits; the goal is to capture data to more accurately assess a driver's risk of collisions and set a rate for the driver.

Privacy

With the continuous collection, analysis, and storage of personal data, there are incumbent risks to consumer privacy. Privacy is the biggest threat to Big Data's expansion and proliferation.

The conundrum with Big Data is that it does not fit into the currently existing privacy legal framework. In particular, Big Data challenges the Fair Information Practices ("FIPs"), which form the basis of all modern privacy law. These practices are based on the principles of notice and consent, as well as awareness and understanding regarding the use of data. These principles are

generally impractical for Big Data, which derives much of the data's value from the secondary and sometimes unimagined use of the data.

Similarly, some legal commentators have disagreed with the proposed international privacy regulations, such as the European Union Data Protection Directive 95/46 EC, because it "relies too heavily on the discredited informed choice model, and therefore fails to fully engage with the impending Big Data tsunami."[18] This is also insufficient considering many consumers fail to read or simply do not understand the length privacy policies and ambiguous terminology contained therein.

Other legal commentators have proposed a regulatory shift from "privacy by consent" to "privacy through accountability."[19] Under this approach, regulators would establish ground rules for the use of data and data users would self-assess their use of the data. Sloppy assessments would subject data users to potential legal liability, regulatory fines, and criminal penalties. This more flexible framework allows for the reuse of data without the constant need for notice and consent.

Some legal commentators have instead proposed empowering individuals with enhanced transparency and access rights.[20] In particular, they propose providing individuals access to their data

---

[18] Ira S. Rubinstein, Big Data: The End of Privacy or a New Beginning? (2012). *New York University Public Law and Legal Theory Working Papers*, Paper 357.

[19] Mayer-Schönberger and Cukier, p. 175.

[20] Omer Tene and Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239 Journal (2013), available at http://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1

in a *usable format* that creates a tangible benefit and value to the individual.[21]

Commentator Andrew Keen, author of *Digital Vertigo*, argues that people want a dark space that protects individual privacy.[22] "The internet needs to learn how to forget. All it knows is how to remember. That's not very human," he said in an interview with CNN.[23] He argues that there will be a backlash against the radical transparency that is idealized in the Big Data movement. By recording everything, we are embracing a culture of digital narcissism, he said.[24] Because the Big Data Revolution is at its beginning, he argues that society should protect privacy with an internet that learns to forget.[25] He uses the example of pollution that occurred during the industrial revolution and compares that to the invasion of privacy that has and will continue to occur in the Big Data revolution until privacy is valued.[26]

European Commission has recognized this issue and proposed new regulations regarding the protection of personal data and a unified "right to be forgotten."[27]

---

[21] *See id.*
[22] Interview with Andrew Keen on CNN, Will We Care About Online Privacy in 20 years?, Stephanie Busari, April 11, 2013, http://www.cnn.com/2013/04/09/tech/privacy-outdated-digital-age/index.html?iref=allsearch.
[23] *Id.*
[24] *See id.*
[25] *See id.*
[26] *See id.*
[27] General Data Protection Regulation, January 25, 2012, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0011:FIN:EN:PDF

While the specific framework for data privacy in Big Data is in flux, there appears to be at least some consensus regarding certain basic principles. Last year, Microsoft invited more than 70 privacy and data protection experts from government, industry, nonprofit organizations, and academia to a global privacy summit in Redmond, Washington. The summit drew participants from 19 countries on five continents, who came together to consider the future of data sources and uses and practical steps to enhance privacy protection.[28]

There was considerable concern at the privacy summit regarding the need for better public awareness and to better inform the public and key privacy stakeholders (e.g. legislators, regulators, press, businesses) about data processing activities as well as the benefits and risks of those activities. Additionally, there was a pressing need for increased standardization, consistency, and interoperability across data protection laws and practices. Finally, there was a widespread understanding about the need to update or enhance the 30-year old Organization for Economic Co-operation and Development ("OECD") Privacy Guidelines, on which most modern data protection laws are based. Recommended updates to these provisions included limits and better transparency in data collection, restrictions for the use of data, security safeguards, redressability, and accountability for data holders.

[28] Notice and Consent in a World of Big Data, November 2012, http://www.microsoft.com/en-us/download/details.aspx?id=35596

Anonymization

Another concern with privacy is the overreliance or overconfidence on anonymization. This privacy protection method involves removing identifiable information so data is used while protecting the anonymity of the data source. Specific methods include redacting, obfuscating, and/or masking data. While this technical approach may work with smaller data sets, it does not effectively protect privacy in Big Data.

For example, computer scientist Dr. Latanya Sweeney illustrated in her PhD thesis at MIT when she used publically available voter registration information to identify Governor William Weld in an anonymized medical database. In fact, Dr. Sweeney found that 87% of the US population can be identified simply using a birth date, zip code, and gender.[29]

In 2006, America Online (AOL) released 20 million search queries posed by users over a three-month period in order to facilitate research on information retrieval. AOL attempted to protect its customers' privacy by anonymizing customer identities and assigning each one a random number. Yet, in only a few hours, two New York Times reporters were able to identify one of the users based on her search history: "landscapers in Lilburn, Ga," several people with last name Arnold, and "numb fingers."[30]

---

[29] Latanya Sweeney, *Simple Demographics Often Identify People Uniquely*. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000

[30] http://www.nytimes.com/2006/08/09/technology/09aol.html?pagewanted=all&_r=0

More recently, the movie rental company Netflix released an alleged anonymized list of 100 million rental records for almost half a million users and incentivized the public to improve its movie recommendation system. Computer scientists were able to identify several customers by linking the Netflix movie ratings to IMDB (an on-line movie ratings website), thereby uncovering potentially sensitive information such as political leanings and sexual orientation. In fact, an in-the-closet lesbian mother sued Netflix for privacy invasion, alleging the company made it possible for her to be 'outed' as gay when it released the allegedly anonymized information.[31]

Thus, companies should scrutinize their anonymization and de-identification procedures to reduce the potential legal, regulatory, and organizational risks.

Trade Secrets and Contractual Agreements

Big Data raises a number of other issues, including proprietary rights in data and the creation of new intellectual property.[32] These issues, however, are not entirely novel and likely the same as or similar to those raised by the Internet, cloud computing, and other network related technologies. Trade secrets in particular, may experience both risks and rewards in the Big Data field.

A trade secret is any information not generally known, that is economically valuable, and subject to reasonable efforts to

[31] http://www.wired.com/threatlevel/2009/12/netflix-privacy-lawsuit/
[32] John Pavolotsky and Jay Seirmarco, Big Data Business Models, New Matter, State Bar of California Intellectual Property Section, Fall 2012, Vol. 37, Issue 3.

maintain its secrecy.[33] Many people think of secret formulas, such as the formulas for Coca-Cola, KFC secret spices, or WD-40. Yet trade secrets can also include a wide variety of technical and nontechnical information. In fact, Big Data may generate new trade secrets for companies. For example, courts have found the following types of information legally protectable as trade secrets: data;[34] databases;[35] compilations of information for storage devices and data modules;[36] generally known elements combined in novel ways;[37] advertising and marketing strategies, plans, and techniques;[38] and customer information.[39] Indeed, companies may find trade secret protection more advantageous to other forms of intellectual property protection.

The heart of the trade secret's value is its secrecy. A trade secret owner must take reasonable efforts to ensure the information's secrecy.[40] He or she must take actual efforts to protect the trade secret so that the trade secret is not

---

[33] *See e.g.* 18 U.S.C. § 1839 (3) (A), (B) (1996); Cal. Civ. Code § 3426.1(d).
[34] *Masonite Corp. v. County of Mendocino Air Quality Mgmt. Dist.*, 42 Cal. App. 4th 436, 449 (1996) (deeming "emission factors" data as trade secret information, as compared to public "emission data").
[35] *MAI Sys. Corp. v. Peak Computer, Inc.*, 991 F.2d 511, 520-23 (9th Cir. 1993).
[36] *Forro Precision, Inc. v. IBM Corp*., 673 F.2d 1045, 1057 (9th Cir. 1982).
[37] *02 Micro Int'l Ltd. V. Monolithic Power Sys.*,420 F. Supp. 2d 1070, 1089 (N.D. Ca. 2006).
[38] *Whyte v. Schlage Lock Co*., 101 Cal. App. 4th 1443, 1455-56 (2002)
[39] *Peerless Oakland Laundry Co. v. Hickman*, 205 Cal. App. 2d 556, 559-60 (1962) (stating that customer likes, dislikes, and needs is "the very kind of information which so often spells the difference between success and failure in a business based on service").
[40] *J. T. Healy & Son, Inc. v. James A. Murphy & Son, Inc.*, 357 Mass. 728, 730-31 (1970).

misappropriated through improper, illegal, or unethical means. The burden is on the trade secret owner to keep the information secret. Furthermore, he or she cannot expect others to hold a higher obligation to keep the information secret.

Big Data may also eliminate trade secret protection from certain confidential information. For example, a company's customer lists may be discovered by a competitor applying Big Data analytics to public shipping manifests and other public delivery information. Indeed, the new Big Data phenomena may threaten the protection of trade secret information that is not actually secret but merely difficult to obtain.

Unlike patent, trademark, or copyright protection, there is no set time period for trade secret protection. A trade secret is protected as long as it is kept secret. However, once a trade secret is lost, it is lost forever.[41] Thus, placing trade secret or confidential information in the hands of a third party for Big Data analytics and other services is a potential risk to maintaining secrecy.

The best way to protect trade secrets and other confidential information, aside from not disclosing them to anyone, is through well-drafted contracts and policies. This includes contracts with both companies providing Big Data services and internal employees who may access the information. First, define the ownership rights in the data. For example, you may want to explicitly state that the Big Data services provider and its employees have no ownership rights in the data or any data

[41] WikiLeaks website publishes classified military documents from Iraq, http://articles.cnn.com/2010-10-22/us/wikileaks.iraq_1_wikileaks-website-classified-documents-iraq-wiki-leaks-iraqis?_s=PM:US.

resulting from their services. The agreement can state that the company and its employees have limited access to the data only for certain reasons. Defining the limits of authorization may also help establish rights under the CFAA if the provider or employee violates the scope of their authorizations, depending on the jurisdiction. Prohibit the unauthorized use or disclosure of company data, including trade secrets and confidential and proprietary information. Contracts may also provide for injunctive relief, liquidated damages, arbitration, and attorneys' fees; companies should seek advice from competent counsel to ensure that any contractual provisions comply with any potential international complications. Special care should be given to address choice of law and forum selection issues which may prove to greatly influence the outcome of any dispute.

Companies should also control access to their data. Agreements with Big Data service providers should restrict the use of their data by outside vendors or third parties. The details of this should be documented in a SLA (service level agreement) and include agreed upon methods for continuous monitoring of the terms/conditions and controls for known risks. Spelling this out in an SLA or a contract is good, but only part of the solution– monitoring for enforcement of the SLA is the key to early detection of issues and the development of a mitigation plan. Provisions may also hold the provider and any subcontractors liable for security breaches.

Contracts should include ongoing confidentiality obligations to protect the company's data in the case of

termination. Additionally, contracts should require the return or deletion of any copies of the data (as appropriate), backup media (e.g. tapes and third party storage) by the service provider or employee after the termination of the agreement. Also consider obtaining a certificate of destruction from independent company if observation of the destruction is not practical by the data owner. Finally, there should be a provision prohibiting the withholding of data by the provider or employee in the case of a dispute.

Companies must be vigilant to ensure that their Big Data policies and agreements, including social networking policies, remain current with changing technology to protect their most valuable assets.

Conclusion

Big Data provides significant benefits for the development and growth of society, including health care, government, education, and various other business sectors. Companies that embrace this new technology and venture into the Big Data realm may distance themselves from their competition but they must be careful and thoughtful. Privacy considerations must be taken into account before data is collected and used in light of the current legal framework. Additionally, companies should scrutinize what they provide to Big Data service providers as well as what data they receive from third parties, how that data will be used, and who else may have access to that data. Well drafted agreements and policies with both Big Data service providers and their employees can help companies ensure that the benefits of Big Data are not outweighed by the potential legal and business risks.