



# The Privacy Implications of Big Data

*June 2013*

Sponsored by:



# The Privacy Implications of Big Data



*This is Big Data, an evolving term but one that is generally described as the ability to capture, store, manage and analyze information that would otherwise be considered too big, too fast or too unstructured for traditional database and analytical software.*

## Executive Summary

Data is continuously generated, and it is being collected on everyone, in every walk of life and in every business sector. From financial transactions to web searches, driving habits to social media posts, the sheer volume of collectable information is staggering. This is “Big Data,” and with the advent of computers able to discover patterns in reams of messy unstructured data, those interested can now gain real-time intelligence about products, services, customer behavior and even the economy. In other words, they can make smarter decisions based on information never available to them before. This ability also raises questions concerning privacy, an issue that has long had both legal and business implications, and one that must be addressed before the power of Big Data can fully be realized.

## Introduction

Take a second to consider your daily data footprint. Think about the websites you visit, queries you make through search engines, the smartphone you carry, credit card transactions you make, toll booths you drive through, GPS signals that direct you and the pictures, blog entries and status updates you post. Think about the vast amount of information about you, your interests, hobbies and routines that are captured and stored on a daily basis. This is Big Data, an evolving term but one that is generally described as the ability to capture, store, manage and analyze information that would otherwise be considered too big, too fast or too unstructured for traditional database and analytical software.

Until recently, the ability to consume this massive amount of information was largely limited to governments, universities and the largest of organizations. Today, however, resources such as commodity hardware, cloud architectures and open source software provide just about any business with the capability to collect, store and process such information.<sup>1</sup> Still in its infancy, the number of organizations currently taking advantage of Big Data is relatively small, but excitement over its possibilities is growing with some predicting its significance to be bigger than the advent of the supercomputer and the Internet.

Advocates of Big Data believe that the ability to collect, store and analyze huge amounts of structured and unstructured data is a breakthrough for businesses and others hungry for real-time intelligence about products, services, customer behavior and even the economy. They believe that with this information, businesses can tailor products and services more precisely to their customers’ needs and expectations, be more competitive in their pricing, and enhance the quality of customer service. By tapping into large-scale sources of information, governments can better allocate scarce resources, healthcare providers can better match patients with treatments, economists can more accurately spot economic trends, and investors may gain new insights into market performance.

The collection and storage of Big Data, however, raises concerns over privacy. In some cases, the uses of Big Data can run afoul of existing privacy laws. In all cases, organizations risk backlash from customers and others who object to how their personal data is collected and used. This can present a challenge for organizations seeking to tap into Big Data’s extraordinary potential, especially in industries with rigorous privacy laws such as financial services and healthcare. Some wonder if these laws, which were not developed with Big Data in mind, sufficiently address both privacy concerns and the need to access large quantities of data to reach the full potential of the new technologies.



*All the personal data collected by companies can cause privacy issues with potential financial and reputational consequences.*

## The Data Explosion

Digital data is continuously collected on almost everyone, and seemingly about everything. To put this into perspective, in just two days as much data is generated as was produced from the dawn of civilization to 2003.<sup>2</sup> This data may be structured – organized to fit into traditional databases – but most of it is unstructured, comprised of such things as text, video, images, and audio recordings which in the past would have been difficult, if not impossible, to use for analytical purposes.

Mass quantities of data are being generated and stored through the personal use of computing and communications devices such as smartphones, tablets, notebooks and PCs. Mobile phones have been a huge contributor to the data explosion: according to the International Telecommunication Union, there were about 6 billion mobile phone subscriptions at the end of 2011, which is equivalent to 87 percent of the world's population. Thirty-two percent of mobile phones sold in 2011 were smartphones.<sup>3</sup>

Businesses also create and store an abundance of data. Some of the data is structured and includes personally identifiable information (PII) such as customer names, addresses, social security numbers, medical records, orders, payment details and birth dates. Some is unstructured and includes information such as GPS output used to track shipments and drivers; operations data from sensors installed in process machinery;<sup>4</sup> email, text and instant messaging between customers, employees, and business partners; social media posts; data exchanges between supply chain partners; and the list goes on and on.

All the personal data collected by companies can cause privacy issues with potential financial and reputational consequences. How this data is protected and used, and what controls people have over their own data, are issues that must be considered and addressed before the expanded use of Big Data analytics can succeed.

## The Untapped Potential in Big Data

Data is constantly collected and stored in data centers around the world, but due to either lack of creativity, technical limitations or concerns around privacy or security, it often is not used productively. Forward-looking governments and businesses, however, increasingly are investing in capabilities to handle and extract value from it. According the Presidential Science Advisor, John Holdren, the US government believes that Big Data holds extraordinary potential for “scientific discovery, environmental and biomedical research, education and national security.” In 2012 the government announced a \$200 million effort by six federal agencies to “uncork” its power.<sup>5</sup>

Within the private sector, examples of how businesses can use – and in some cases already are using – Big Data include:

- New product development.
- Identifying cross-selling opportunities among products.
- Tracking customer attitudes towards particular companies, brands and products.
- Identifying which customers are most likely to leave for a competing vendor.
- Detecting fraud.

McKinsey Global Institute has identified various ways that Big Data can create value for and change the way businesses are designed, organized and managed. According to McKinsey, the benefits of Big Data include:<sup>6</sup>

- More effective business models.
- Greater innovation in products and services.
- Products tailored to refined customer segments.
- Faster time to market for new products.
- Better decision-making.
- Improved organizational performance.



*In what is sometimes referred to as the “privacy bargain,” a trade-off is often required between the convenience provided by technology and privacy.*

Companies that collect large quantities of customer data during the course of business such as insurance carriers, retailers, transportation companies and communications providers are in the best position to utilize Big Data tools and techniques to transform processes, alter corporate structures and facilitate innovation.<sup>7</sup> This should provide a significant competitive advantage, one that over the next few years will allow them to outperform their peers by 20 percent, by one estimate.<sup>8</sup> Even companies without a large volume of their own data may be able to benefit from Big Data using data available from information brokers and various public sources.

Living in a data driven society requires adherence to the laws that protect an individual's right to privacy. Advancements in technology, however, occur at a rapid pace and create privacy issues that were not contemplated when the laws were drafted. As a result, some question whether current privacy laws are appropriate – or even applicable – to the Big Data world. Outmoded privacy laws, however, are only one of the privacy-related obstacles that stand in the way of realizing Big Data's full potential. Beyond the legal framework lies a thicket of social, business and ethical issues concerning the use of private data that must yet be addressed.

A recent incident highlights concerns over how personal information is used for Big Data applications. Privacy advocates are in a stir about revelations made by a whistleblower regarding the U.S. government's use of Big Data to essentially spy on its own citizens. According to the whistleblower, the National Security Agency (NSA) has been collecting data on millions of American phone customers. The data included subscriber data, recipients, locations, times and durations of calls. The government claims the collection of this information is for national security purposes and plans to continue with the program but privacy advocates object to the practice and have responded with lawsuits.

## Privacy and the Legal Ramifications of Big Data

In what is sometimes referred to as the “privacy bargain,” a trade-off is often required between the convenience provided by technology and privacy. This trade-off frequently occurs with Big Data, as is the case with electronic toll collection systems that allow us to breeze through toll gates, but leaves behind information about our travels. The trade-off is especially evident in the context of the Internet. As people conduct their lives online, data is captured and stored about websites visited, purchases made, queries searched, messages and images posted and social media status' updated among countless other activities.

Concerns about the privacy implications of technology are nothing new, with Big Data being only the latest development to keep the issue at the forefront of public interest and conversation. Many of the privacy issues are much the same as those grappled with over the past several decades, but they are heightened by the ability to access and analyze much more personal information than ever before and, because of the nature of Big Data, the fact that it is often impossible to know who has what data and where it is physically located. People are often uncomfortable with the fact that the fluidity of data in a Big Data world and the frequent lack of transparency can make it difficult to hold a business, government or an individual accountable for how personal data is used or misused. Maintaining a balance between the benefits of technology and privacy is an emotional issue of high interest to many people.



*Healthcare is an industry that collects significant PII, is highly regulated as concerns privacy, and also has begun to embrace Big Data. Big Data holds enormous potential for improving the cost, quality and safety of patient care.*

The government, both at the federal and state level, has long regulated the privacy of personal information. For example, The Fair Credit Reporting Act, passed in 1970, permitted credit bureaus to collect personal financial information but restricted how that information could be used. In the US however, there is no comprehensive piece of legislation that broadly covers the issue of privacy. Instead, at the federal level, there is a patchwork of more than 20 laws with privacy provisions. Some of the most stringent laws are designed for specific industries or populations.

Examples include:

- Health Insurance Portability & Accountability Act (HIPAA)
- Health Information Technology for Economic & Clinical Health Act (HITECH)
- Gramm-Leach Bliley
- Children's Online Privacy Protection Act (COPPA)
- USA Patriot Act
- Fair Credit Reporting Act (FCRA)

Every industry does not require the same level of collection and use of personal identifiable information (PII). Consequently, the laws listed above primarily focus on three areas, healthcare, financial services and children. Healthcare is an industry that collects significant PII, is highly regulated as concerns privacy, and also has begun to embrace Big Data. Big Data holds enormous potential for improving the cost, quality and safety of patient care. For this reason, an analysis of the healthcare industry provides an especially revealing case study on the tensions between the benefits of Big Data and the need to protect patient privacy.

## Healthcare

Over the past decade the healthcare industry has made significant strides in the broad based implementation of electronic medical records. Many are just now getting used to using these digitized systems. Others – mostly the larger medical and healthcare systems – have already taken the next leap in attempting extract value from it. Big Data analytics is allowing these organizations to analyze huge amounts of patient data, perform predictive modeling, and identify clinical risk factors among countless other benefits.<sup>9</sup> “Although these efforts are still in their early stages, they could collectively help the industry address problems related to variability in healthcare quality and escalating healthcare spend” according to a January 2013 report from McKinsey & Company. In fact, according to the same report, researchers say that Big Data could reduce healthcare spending by anywhere from \$300 - \$450 billion, or 12 to 17 percent of the total US healthcare costs.<sup>10</sup>

If these predictions are true, the benefits of Big Data seem clear. Privacy advocates see it a bit differently though. They see the use of Big Data coming at the expense of patients, patients who they believe at the moment have no control over what information about them is collected and how it is being used. They argue that patients should be able to know exactly how the industry obtains this enormous amount of information and who they are obtaining it from. They feel that collecting and analyzing Big Data without informed consent is illegal and in violation of already established privacy laws such as HIPAA and HITECH.

Other concerns revolve around the security – or lack thereof – of protected health information (PHI). The healthcare industry as a whole has had an abysmal record of securing its structured data (i.e. patient records). According to the Ponemon Institute/Experian Data Breach Resolution Report, “although organizations have a clear understanding of the risks and potential consequences of data breaches, many are not taking adequate steps to protect themselves.”



*Plaintiffs' attorneys are aggressively probing the boundaries of existing federal privacy laws – laws which clearly did not contemplate Big Data when they were written – and increasingly they are proposing novel new theories of liability.*

This leaves privacy advocates wondering how an industry that has traditionally struggled with protecting its data is going to protect the terabytes of newly generated electronic health records plus new sources of unstructured data it now collects.

Some privacy advocates question whether the current healthcare privacy laws are sufficient for the evolving Health Information Technology (HIT) environment. While some suggest that it is less about the laws' limitations and more about their enforcement, others believe current laws are wholly inadequate for the Big Data world. An increased level of transparency about what data is collected, how it is used, how it is protected, and the potential benefits it brings is needed to bridge this disconnect. A balance must be found in order for the power of Big Data to be fully realized, but not at the expense of an individual's privacy.

## Other Uses of Big Data

Federal privacy laws govern the use of medical and personal financial information, but as just discussed, how these laws apply to Big Data is not always clear. Even greater ambiguity surrounds privacy issues in other areas, such as the use of data concerning goods purchased, websites visited, and smartphone location trails. Plaintiffs' attorneys are aggressively probing the boundaries of existing federal privacy laws – laws which clearly did not contemplate Big Data when they were written – and increasingly they are proposing novel new theories of liability. Additionally, allegations of state consumer protection law violations also are often tacked onto complaints alleging privacy law violations.

A recent lawsuit against Google by Android users, for example, alleges the company violated the federal Computer Fraud and Abuse Act as well as California's Unfair Competition Law by transmitting geolocation data and other personal information to app developers. Similarly, in the largest privacy class action suit ever, plaintiffs allege the software used by a company that collects and sells consumer information violates the Stored Communications Act, the Electronic Communications Privacy Act, and the Computer Fraud and Abuse Act, plus the Illinois Consumer Fraud and Deceptive Practices Act.

Regardless of whether a legal framework exists for privacy protection, organizations should be concerned about backlash from customers who displeased about how their data is being treated. Privacy experts note that customers are most unhappy when data is collected secretly, when they don't understand how it is being used, and when they are unclear as to how they benefit from it.<sup>11</sup>

Privacy advocates claim that it is in an organization's best interest to provide full disclosure, both to cultivate goodwill and to avoid creating incentives for restrictive and burdensome privacy legislation. Many organizations, however, presently do not do a good job of making their data-usage intentions clear. Principles for responsible and transparent data management have been proposed by a number of organizations.

Privacy advocates also encourage data collectors to make it easier for customers to opt in or out of having their information used. Technology solutions are being introduced to enable individuals to specify how their data is allowed to be used.



## Conclusion

Many people regard privacy as a fundamental right. Failure by companies, researchers and government agencies to respect that right will create pressure for new solutions that may keep Big Data from reaching its full potential. In a recent whitepaper addressed to policymakers, the Software & Information Institute of America (SIIA) warns that “Policies that seek to curb the use of data could stifle this nascent technological and economic revolution before it can truly take hold.”<sup>12</sup>

The current patchwork of federal privacy laws is proving inadequate to address privacy challenges in the world of Big Data. Almost certainly, more laws are coming – privacy laws have been introduced in Congress from time to time, and the Obama administration and the Federal Trade Commission continue to press Congress to enact online privacy legislation. But many experts acknowledge that laws alone will not guarantee that privacy will be fully safeguarded. “Policy frameworks that constrain how data can be linked, shared and used ... are increasingly less effective and anachronistic in today’s hyperconnected world,” notes a report from the World Economic Forum (WEF).<sup>13</sup>

The onus is on organizations to behave responsibly. According to the SIIA: “For DDI [data-driven innovation] to reach its full potential, it must be built on a foundation of good data stewardship and trust. Transparency is the critical first step.”<sup>14</sup> Companies increasingly recognize the need to address privacy concerns to maintain the flow of useful data.

Technology may provide solutions for empowering people to control how their data is used. “Privacy by Design” – baking privacy into a company’s products rather than relying exclusively on written privacy policies – has caught the attention of policymakers and privacy regulators, and is being implemented at a growing number of companies. The authors of the WEF report encourage organizations to create “simple-to-use tools that encourage individuals to become engaged in setting the policy governing use of data and to be able to change those settings over time without being overwhelmed.”<sup>15</sup> They envision a future in which collected data is tagged with software code that includes each individual’s preferences for how his or her data is used.

For the full potential of Big Data to be harnessed, companies must not only adhere to the current patchwork of federal and state privacy laws, they must take steps to build trust by demonstrating responsible data stewardship and engaging customers in controlling the use of personal data. The opportunities presented by Big Data are staggering, but if privacy concerns are not adequately addressed, its benefits may never be fully realized.

### NOTES:

<sup>1</sup> Ed Dumbil, O’Reilly Radar, “What is Big Data?: An introduction to the big data landscape”, (January 2012), <http://radar.oreilly.com/2012/01/what-is-big-data.html>

<sup>2</sup> Chris Forsyth, 4syth.com, “For Big Data Analytics There’s No Such Thing as Too Big: The Compelling Economics and Technology of Big Data Computing”, (March 2012)

<sup>3</sup> mobiThinking, “Global mobile statistics 2012 Part A: Mobile Subscribers; handset market share; mobile operators”, (2012), <http://mobiThinking.com/mobile-marketing-tools/latest-mobile-stats/a#subscribers>

<sup>4</sup> McKinsey & Company, “Big data: the next frontier for innovation, competition and productivity”, (May 2011), [http://www.mckinsey.com/insights/mgi/research/technology\\_and\\_innovation/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation)

<sup>5</sup> USAToday.com, “‘Big Data’ disguises digital doubts”, (6/16/2012)

<sup>6</sup> McKinsey & Company, “Big data: the next frontier for innovation, competition and productivity”, (May 2011), [http://www.mckinsey.com/insights/mgi/research/technology\\_and\\_innovation/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation)

<sup>7</sup> McKinsey & Company, “Big data: the next frontier for innovation, competition and productivity”, (May 2011), [http://www.mckinsey.com/insights/mgi/research/technology\\_and\\_innovation/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation)

<sup>8</sup> Chris Forsyth, 4syth.com, “For Big Data Analytics There’s No Such Thing as Too Big: The Compelling Economics and Technology of Big Data Computing”, (March 2012)

<sup>9</sup> Elizabeth S. Roop, “Big Data Creates Big Privacy Concerns,” *For the Record*, Vol. 24 No. 16 P. 10, (September 10, 2012), <http://www.fortherecordmag.com/archives/091012p10.shtml>

<sup>10</sup> Peter Groves, Base Kayyali, David Knott, Steve Van Kuiken, “The ‘big data’ revolution in healthcare”, *McKinsey & Company*, (January 2013)

<sup>11</sup> “Privacy Concerns: Big Data’s Biggest Barrier?” *InformationWeek* <http://www.informationweek.com/big-data/news/big-data-analytics/privacy-concerns-big-datas-biggest-barrier/240156861>

<sup>12</sup> *Data-Driven Innovation: A Guide for Policymakers*, SIIA p.4

<sup>13</sup> *Unlocking the Value of Personal Data: From Collection to Usage*, World Economic Forum, February 2013 p. 4 [http://www3.weforum.org/docs/WEF\\_IT\\_UnlockingValuePersonalData\\_CollectionUsage\\_Report\\_2013.pdf](http://www3.weforum.org/docs/WEF_IT_UnlockingValuePersonalData_CollectionUsage_Report_2013.pdf)

<sup>14</sup> SIIA p. 4

<sup>15</sup> WEF p. 18

