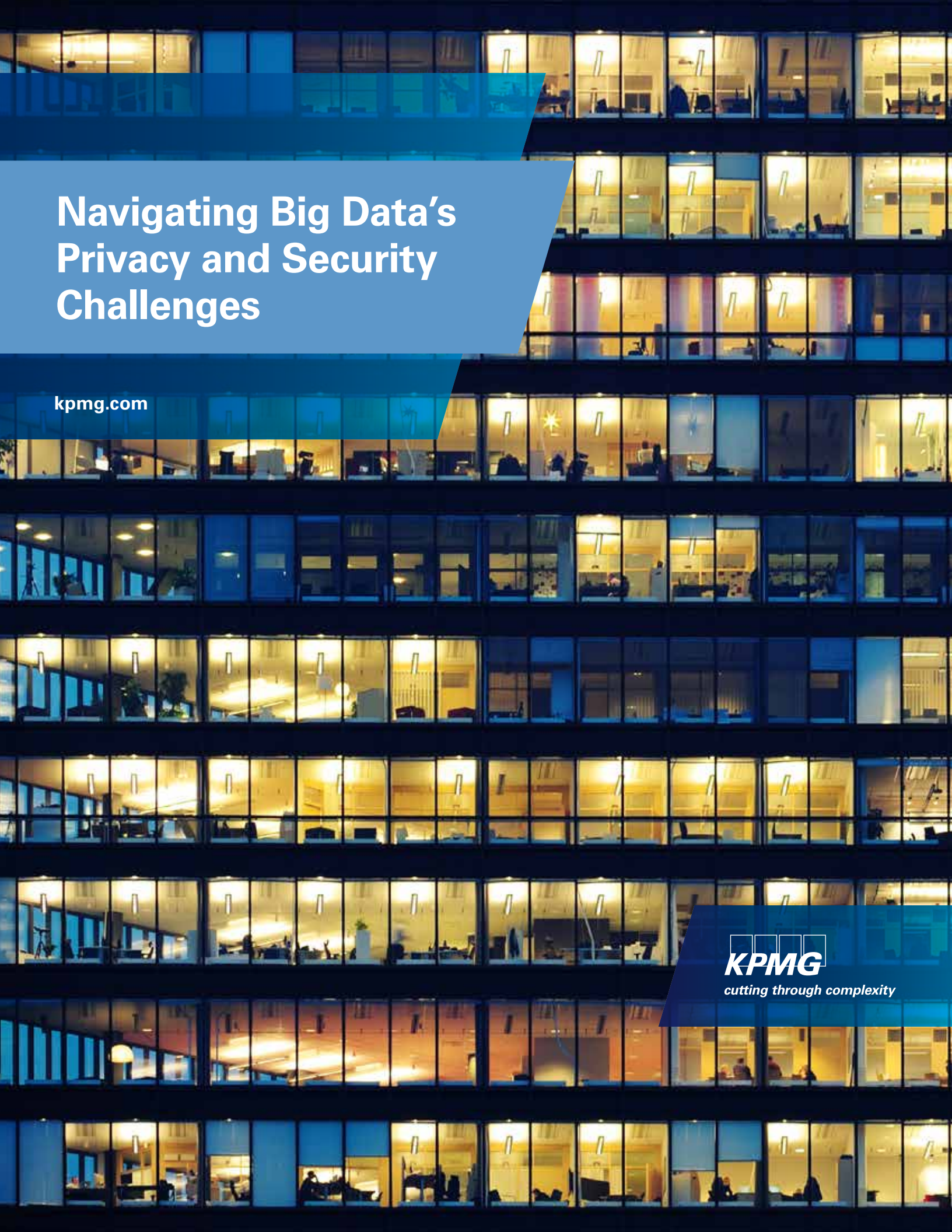


Navigating Big Data's Privacy and Security Challenges

kpmg.com

KPMG

cutting through complexity

About the authors

Greg Bell is a principal in the Atlanta office of KPMG LLP's (KPMG) Advisory Services Practice, and serves as KPMG's National Practice Leader for our Information Protection and Business Resilience (Security, Privacy, and Continuity) practice. Greg is experienced with various areas of Information Management and Information Security, with particular specialization in the fields of IT risk management and business enablement. He has extensive knowledge and experience managing complex projects implementing, administrating, and securing complex client-server and heterogeneous network technologies.

Doron Rotman is the National Privacy Service Leader for KPMG based out of the Santa Clara, CA office. Doron brings more than 25 years of industry experience, and he has led many of the firm's largest privacy and security engagements, serving complex, global organizations. Doron has presented at various conferences on Big Data Privacy and Privacy by Design, most recently at ISACA NA CACS 2014.

Mike VanDenBerg is a Director in the Dallas, TX office of KPMG's Information Protection and Business Resilience practice and has 11 years of experience within Information Security and Data Privacy. Mike has broad experience in the Big Data Privacy space and has become a go-to leader in the firm for Big Data Privacy engagements. Mike's Big Data experience includes benchmarking and strategy development in an era where Big Data innovation outpaces regulation.

KPMG would also like to thank contributors
Chris Koehnecke, Manager, Advisory;
Chris Kypreos, Senior Associate, Advisory;
and Sarah Pipes, Senior Associate, Advisory.

Big Data is a transformative, pervasive avalanche that is not going away and just keeps accelerating. Organizations are rapidly implementing Big Data programs to strategically change their organizational business models to gain a competitive advantage, increase their bottom line, and expand their global presence. However, as such programs grow they face potential conflict with an increasing number of international laws and standards. Therefore, organizations must seek an appropriate balance of opportunities and challenges as they build out their Big Data governance programs to optimize Big Data's benefits, while properly addressing issues related to global privacy, security, and compliance.

The "four V's" of **Big Data**

To put the acceleration of Big Data in perspective, the data sets are increasing to sizes that were unfathomable nearly 20 years ago and continue to grow. It was reported in 2012, for example, that 90% of all available data had been created in the previous two years.[1] With this constantly expanding range, Big Data sets are more commonly understood by their four characteristics: ***Volume, Velocity, Value,*** and ***Veracity.***[2]

Furthermore, Big Data sets are either too large or too fast-changing to be analyzed using traditional relational and multidimensional database techniques or commonly used software tools to capture, manage, and process the data in a reasonable elapsed time.[3]

Big Data is transforming **major industries**

Big Data is not confined to a single industry. KPMG LLP's (KPMG) clients, particularly in the public utilities, telecommunications, healthcare, and financial fields, all recognize Big Data's value while they attempt to harness its characteristics.

Utility companies, for example, have implemented SmartGrid technology throughout the United States and Europe, translating into improvements in Outage Response, Renewable Reliability, and Load Forecasting.[4] Medical companies are using Big Data and analytics to improve customer treatment options.[5, 6] Financial institutions are using Big Data to better identify their customer markets.[7]

As Big Data continues to transform industries, organizations face a growing demand for talent to help them take advantage of Big Data's opportunities.[8] It is clear that "organizations which benefit from Big Data & Analytics will have a competitive edge when it comes to better and faster business decisions."[9]

Big Data: **Big risks**

Although the opportunities are available for those who take advantage of them, Big Data carries significant security, privacy, and transfer risks that are real and will continue to escalate. It is important that companies give consideration to

---

1 Johnson, Jeanne, "Big Data + Big Analytics = Big Opportunity" (Financial Executive July/August 2012)

2 Kobielus, James, "Measuring the Business Value of Big Data." The Big Data & Analytics Hub, http://www.ibmbigdatahub.com/blog/measuring-business-value-big-data.

3 "Big Data: Impacts & Benefits." An ISACA White Paper (March 2013).

4 McMahon, Jeff, "Big Data From Smart Grid Tells Utilities More Than They Want to Know (Forbes) (Sept. 26, 2013).

5 Versel, Neil, "How Hospitals are Dealing with Big Data." (US News) (Oct. 15, 2013)

6 "WebChartMD's Big Data Tool Expands Use of Unstructured Data in Healthcare" (PRNewswire) (Oct. 15, 2013).

7 Hickins, Michael, "Banks Using Big Data to Discover 'New Silk Roads.'" (Wall Street Journal) (Feb. 6, 2013).

8 Manyika, James, Chui, Michael, et al. "Big Data: The next frontier for innovation, competition, and productivity." (McKinsey Global Institute) (May 2011).

9 "Big Data & Analytics: turning conceptual thinking into powerful results." KPMG NL.

the risks related to identification,[10] re-identification,[11] predictive analysis,[12] the indiscriminate collection of data,[13] and increased risk of data breach, which can result in new data creation when combining data from a multitude of sources[14] as organizations seek to optimize their Big Data programs.

Some examples of these sources: Law enforcement has turned to Big Data surveillance in an attempt to fight crime.[15] Social media collects mass amounts of data creating larger, searchable digital footprints.[16] Internet providers offer a multitude of integrated services that track user information and predict actions and interests.[17]

In January 2014, President Barack Obama commissioned a report to examine the Big Data opportunities while preserving interests related to privacy. The White House review identified a series of Harms and Policy Recommendations associated with Big Data, including concerns associated with discrimination:

> *An important finding of this review is that while Big Data can be used for great social good, it can also be used in ways that penetrate social harm and render outcomes that have inequitable impacts, even when discrimination is not intended. Small biases have the potential to become cumulative, affecting a wide range of outcomes for certain disadvantaged groups. Society must take steps to guard against these potential harms by ensuring power is appropriately balanced between individuals and institutions.*[18]

As organizations seek to optimize their Big Data Programs, they must remain cognizant of Big Data's risks.

Regulatory bodies are now aware of the risks Big Data poses and recent efforts to police the Big Data space, demonstrate that companies cannot operate in a vacuum ignoring the consequences of their actions.

---

[10] Sweeney, Latanya, "Simple Demographics Often Identify People Uniquely." Carnegie Mellon University.

[11] Buytendjik, Frank and Heiser, Jay, "Privacy and Ethical Concerns Can Make Big Data Analytics a Big Risk Too." (Gartner) (March 5, 2013).

[12] Crawford, Kate and Schultz, Jason, "Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms." (Boston College Law Rev. Vol. 55. No. 1 2014) (Oct. 1, 2013).

[13] Ramirez, Edith, FTC Chairwoman, Technology Policy Institute Aspen Forum (Aug 19, 2013).

[14] Solove, Daniel, "Introduction: Privacy Self-Management and the Consent Dilemma."

[15] Sengupta, Somini, "Privacy Fears Grow as Cities Increase Surveillance." New York Times (Oct 16, 2013).

[16] Hartzog, Woodrow and Sellinger, Evan, "The Chilling Implications of Democratizing Big Data: Facebook Graph Search is only the Beginning." (Forbes) (Oct. 16, 2013).

[17] "Privacy & Big Data." An ISACA White Paper (Aug 2013).

[18] "BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES", Executive Office of the President (May 2014).

The existing **regulatory landscape**

Regulatory agencies and industry standards have proven that they will hold companies responsible and accountable for their actions as "the risk of consumer injury increases as the volume and sensitivity of data grows."[19]

Federal Trade Commission (FTC) Chairwoman Edith Ramirez addressed Big Data from the regulatory perspective: *"The fact that 'Big Data' may be transformative does not mean that the challenges it poses are, as some claim, novel or beyond the ability of our legal institutions to respond. …. The challenges it poses to consumer privacy are familiar, even though they may be of a magnitude we have yet to see. The solutions are also familiar. And, with the advent of Big Data, they are now more important than ever."*

Similar to U.S. privacy regulations, Big Data is generally dominated by sectoral privacy laws. The United States does not have a national privacy law, or laws specific to Big Data; however, there are existing laws restricting the collection, use, and storage of specific personal information types including financial, health, and children's data. In some cases, these laws have been updated to respond to collection practices made possible by new technology, namely, data-gathering tools such as social media and mobile applications.[20]

It is possible to assess previously passed Big Data-related regulations to highlight high-risk or sensitive Big Data impact areas. For example, the provision in the Fair Credit Reporting Act (FCRA) that requires that individuals be notified of negative decisions made using databases, highlights the fact that negative decisions carry greater risks than positive ones. The Children's Online Privacy Protection Act (COPPA) requires parental consent prior to the collection of the information of minors. It reflects the need for added layers of consent prior to disclosing and use sensitive information.

More broadly, Section 5 of the FTC Act requires the FTC to prosecute unfair or deceptive acts or practices that may affect interstate commerce, and to prevent "unfair" commercial practices, but these are not narrowly defined in the Big Data context. Likewise, best practices exist (e.g., Generally Accepted Privacy Principles), but those tailored to Big Data are difficult to obtain at this point as they are not standardized or widely publicized beyond the implementing organizations. Benchmarking existing practices gives insight into strong implemented or planned processes, but not many such studies are made publicly available.

"Organizations learn by doing, but mistakes can lead to severe reputational impact, even regulatory action."[21] This poses the ethical dilemma: How is a responsible but innovative company expected to behave?

The European regulatory privacy landscape is currently evolving as the European Commission is in the process of implementing Data Protection reform to replace the existing EU Data Protection Directive. The proposed Regulation contains clauses that present potential challenges to the use of Big Data including guaranteeing data subjects a "Right to be Forgotten" and more options for Explicit Consent.[22]

In fact, the Court of Justice of the European Union issued a ruling on May 13, 2014 requiring Google to remove from its search results personal data related to a Spanish man contained in a 1998 news article. The Court further held to whether the directive enables data subjects to a "right to be forgotten":

> *If it is found, following a request by the data subject, that the inclusion of those links in the list is, at this point in time, incompatible with the directive, the links and information in the list of results must be erased.*[23]

As a result, initiatives including the Article 29 Working Party's Opinion on Purpose Limitation[24] and the Big Data Public Private Forum (BIG)[25] seek to provide clear strategic guidance to the growth of Big Data across Europe.

While Big Data concerns stem from familiar complexities, they also face new issues. The prime challenge with Big Data is compliance with these existing laws when the data complexity has increased multifold. The security component of privacy also faces additional risks, because the threat of a breach or other security issue increases in value and risk when the data in question is of Big Data scale. Regulators have recently demonstrated their enforcement powers levying large fines in high-profile cases, including the largest settlement ever under HIPAA ($4.8 Million)[26], and industry speculation that the Target breach could result in fines between $400 million and $1.1 billion.[27]

---

[19] Supra note 13.

[20] The FTC revised the COPPA requirements in 2013 to include Third Parties such as advertising networks and covers unique IDs in cookies, along with IP addresses and device identifiers, which can "recognize a user over time and across different sites or online services."

[21] *Supra* fn 11

[22] Progress on EU data protection reform now irreversible following European Parliament vote (http://europa.eu/rapid/press-release_MEMO-14-186_en.htm)

[23] Court of Justice of the European Union PRESS RELEASE No 70/14, "An internet search engine operator is responsible for the processing that it carries out of personal data which appear on web pages published by third parties" (May 13, 2014)

[24] Article 29 Working Party Clarifies Purpose Limitation Principle; Opines on Big and Open Data (https://www.huntonprivacyblog.com/2013/04/articles/article-29-working-party-clarifies-purpose-limitation-principle-opines-on-big-and-open-data/)

[25] http://news.theentrepreneurshow.com/?p=2974

[26] Data breach results in $4.8 million HIPAA settlements, Department of Health and Human Services (May 7, 2014) (http://www.hhs.gov/news/press/2014pres/05/20140507b.html)

[27] Target could face $1.1 billion in fines: Analyst, CNBC (Jan 30, 2014) (http://www.cnbc.com/id/101378232)

Five Key Big Data Security **and Privacy Challenges**

KPMG has identified five key security and privacy challenges organizations must address to help ensure proper control of their Big Data program:

1 ***Big Data governance***

The implementation of Big Data initiatives may lead to the creation or discovery of previously secret or sensitive information through the combination of different data sets. Organizations that attempt to implement Big Data initiatives without a strong governance regime in place, risk placing themselves in ethical dilemmas without set processes or guidelines to follow. Therefore, a strong ethical code, along with process, training, people, and metrics, is imperative to govern what organizations can do within a Big Data program.

2 ***Maintaining original privacy and security requirements (original intent) of data throughout the information life cycle***

Data that is collected and used for Big Data will likely be correlated with other data sets that may ultimately create new data sets or alter the original data in different, often unforeseen ways. Organizations must make sure that all security and privacy requirements that are applied to their original data sets are tracked and maintained across Big Data processes throughout the information life cycle from data collection to disclosure or retention/destruction.

3 ***Re-identification risk***

Data that has been processed, enhanced, or changed by Big Data programs may have benefits both internal and external to the organization. Often, the data must be anonymized to protect the privacy of the original data source, such as customers or vendors. Data that is not properly anonymized prior to external release (or in some cases, internal as well) may result in the compromise of data privacy as the data is combined with previously collected, complex data sets including geo-location, image recognition, and behavioral tracking. If data is simply de-identified, possible correlation between data subjects contained within separate data sets must be evaluated, as third parties with access to several data sets may be able to reidentify otherwise anonymous individuals.

4 ***Third parties – usage and honoring contractual obligations***

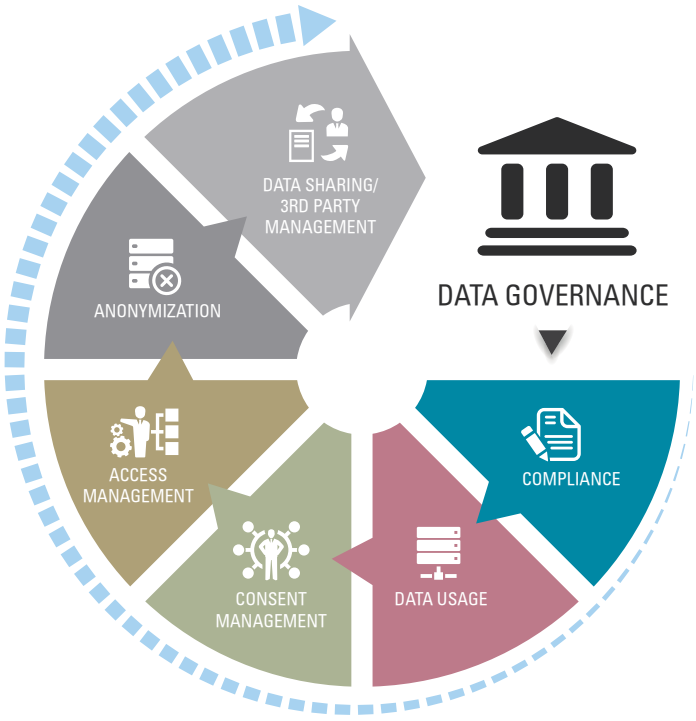
Matching data sets from other organizations may unlock insights using Big Data that an organization could not uncover with its data alone. It may also pose significant risk, as the security and privacy data protections in place at the third-party organization may not be adequate. Prior to sharing data with third parties, organizations must evaluate their relevant practices and decide whether they are satisfactory.

5 ***Interpreting current regulations and anticipating future regulations***

As noted, the United States and the EU do not have laws or regulations specific to Big Data; however, there are existing laws restricting the collection, use, and storage of specific personal information types, including financial, health, and children's data. Additionally, Big Data compliance has seen an increased degree of regulatory scrutiny, as evidenced by the FTC's recent emphasis on Data Brokers and the Article 29 Data Protection Working Party's Opinion on the potential impact of the purpose limitation principle on Big Data and open data. To keep current with quickly changing and new implemented laws, companies must perform an initial inventory of applicable laws and update this inventory on a regular basis.

KPMG's approach to **Big Data security and privacy**

KPMG has developed an approach that identifies specific focus areas that companies should take into account while building out a Big Data Privacy and Security program. Our approach relies on the standard Information Governance life cycle to assist organizations as they appropriately minimize their risks and maximize control over Big Data.



1. **Data governance and retention**
   As the foundation for any Big Data program, a data governance program must be established that provides clear direction for how the data is handled and protected by the organization. This program begins with a clear organizational structure around data governance *(Who owns the data? Who is responsible for protecting the data?)*, followed by additional key components such as policies, standards, and procedures including data monitoring and data retention.

2. **Compliance**
   Organizations must identify and understand the security and privacy regulations that apply to the data they store, process, and transmit. Similarly, they are also responsible for compliance with the contractual provisions contained within their agreements with third parties and other service providers, as well as their own privacy policy. Therefore, it is essential that organizations establish a Big Data compliance program that provides the necessary oversight to monitor compliance with their regulatory and contractual commitments.

   Compliance requires developing a comprehensive control framework and risk-based road map for implementation. Companies can take advantage of automated controls and transition from manual efforts to ensure ongoing compliance.

3. **Data use cases and data feed approval**
   Organizations must manage their Big Data usage through the identification of potential use cases for the data. Once an organization understands the potential use cases, it can mature its Big Data program through the implementation of a formal use-case approval process, which includes formal risk assessments prior to the adoption of new data feeds.

A key consideration in the adoption of any new data feed is that the potential risk for re-identification increases when existing data feeds are combined with new data feeds.

4. **Consent management**
   Customer consent management is critical to the success implementation of any Big Data governance regime. Customer consent requires the following components:

   i. *Transparency* – Organizations should provide its customers with a clear understanding of the information the organization collects and how the information will be used.

   ii. *Consistency* – Organizations should provide consistent consent mechanisms across all products, and capture Big Data preferences up front.

   iii. *Granularity* – Organizations should allow customers to provide or withdraw their consent at the individual device level and not at a larger account level.

5. **Access management**
   Given the amplified size and scope of Big Data, organizations must effectively control who within the organization has access to the data sets. This requires a comprehensive access management regime including review and approval of new user access requests and periodic reviews of existing user access to ensure privilege requirements meet security and compliance requirements. Finally, organizations should adopt segregation of duties where access to systems is based on job function.

   Organizations can automate the process by leveraging policy engines or access management tools to implement Attribute Based Access Controls (ABAC). This will help them make dynamic access decisions and integrate with existing tools and directories for provisioning and certification.

6. **Anonymization**
   Anonymization means removing all Personally Identifiable Information (PII) from a data set and permanently turning it into non-identifying data. Data anonymization is critical to the long-term use of Big Data while maintaining the privacy of the original sets. While the concept of anonymization is complex and few companies reveal how they achieve full anonymization, organizations must take appropriate measures to avoid data re-identification. This requires monitoring of anonymization requirements and analyzing the risks of re-identification prior to the implementation of a particular anonymization technique, including information correlation across multiple data sets.

7. **Data sharing/third-party management**
   Big Data concerns amplify exponentially as the data is combined with additional data sets. Organizations maintain a responsibility to their customers as they share data with third parties. Effective third-party management requires the inclusion of specific Big Data provisions within contractual agreements. Additionally, organizations should align Big Data with its overall strategy for the performance of third-party assessment to ensure ongoing monitoring of third parties for compliance with data-sharing agreements.

The big picture: **Big Data**

Big Data presents myriad challenges and opportunities for decision makers, and has already proven itself at corporations, nonprofits, law firms, government agencies, and other entities that span the globe. Security, compliance, and regulatory concerns of Big Data vary greatly by countries and other jurisdictions. As the utilization increases along with the rising rate of productive commercial activity, so will the benefits for those who move forcefully to embrace and maximize the opportunities of Big Data, while proactively minimizing data sharing, privacy, and transfer risks. In conclusion, KPMG urges companies to understand that:

- Implementing an enterprise-wide Big Data Governance Program with cross-department oversight is crucial to taking advantage of Big Data's true potential.
- Privacy is not just a "nice to have" – Big Data privacy and security risks are real and companies must understand these risks. Companies can no longer collect unimaginable volumes of data without consequence. Discriminatory analysis and data breaches are tangible risks that any company must be aware of while mobilizing their Big Data programs.
- Regulatory bodies will enforce and protect privacy and security in the Big Data age – regulators understand Big Data privacy and security risks and have shifted their enforcement focus on protecting consumers.

Case Study: **Boosting Big Data with benchmarking**

Client: Global Telecom Company

Client challenge: Client sought guidance on industry practices for privacy and security as it began to operationalize its Big Data Program

Summary of work: KPMG used its Big Data Privacy and Security methodology to benchmark the client's existing Big Data Program capabilities. KPMG's assessment included both a public and private benchmarking. The public benchmarking entailed research and compilation of publicly available information on Big Data privacy and security practices, policies, and procedures. To conduct the private benchmarking, KPMG submitted a questionnaire to external enterprises comparable to the client to obtain a high-level overview of their Big Data organizations. At the end of the assessment, KPMG compared results from its Big Data research with the client's Big Data initiative to assist in prioritizing recommendations to enhance or extend the existing information governance strategy and related privacy framework.

KPMG's value proposition: Your next step in harnessing the potential of Big Data is with KPMG. Whether your organization is in the planning stages or has a mature Big Data program, KPMG can assist in navigating the Privacy and Security challenges posed across all phases of the life cycle. KPMG takes a multidisciplinary approach where different groups in the firm specialize in distinct areas across the Big Data life cycle and collaborate to provide insight to your organization. A sampling of our previous engagements includes:

- Big Data Analytics Enhancement: Provide additional capabilities and resources to your analytics platform to improve your Big Data insight development and program development.
- Big Data Security and Privacy Program Assessments: Define and document your current-state Big Data initiatives to identify the maturity of the program in all facets of the life cycle and identify gaps and opportunities for improvement.
- Big Data Security and Privacy Program Development: Develop a future state road map to enable your organization's Big Data goals with critical security and privacy requirements.
- Third-Party/Vendor Assessments: Assess third parties and vendors for compliance with contractual and regulatory requirements for Big Data.
- Big Data Analytics Platform Assessment and Development: Assess your currently Big Data analytics capabilities and identify opportunities for development and improvement.

Glossary

ABAC – Attribute Based Access Controls

Anonymization – removing all Personally Identifiable Information from a data set and permanently turning it into nonidentifying data.

Children's Online Privacy Protection Act (COPPA) – federal privacy law that applies to the online collection of personal information by persons or entities under U.S. jurisdiction from children under 13 years of age.

Data Governance – the management and protection of company information assets throughout the information life cycle. Components include privacy, Information life cycle management (ILM), data classification and data-flow analysis.

Fair Credit Reporting Act (FCRA) – federal law that regulates the permissible collection, dissemination, and use of consumer information by consumer reporting agencies.

Information Life cycle – the full cycle of data within an enterprise that commences with collection, storage, usage, transfer, and destruction.

Personally Identifiable Information (PII) – information that can be used on its own or with other information to identify, contact, or locate a single person, or to identify an individual in context.

Predictive Analysis – a variety of statistical modeling and data mining that analyzes current and historical facts to make predictions about future events.

Sensitive Data – a subset of personal information that is subject to a higher level of privacy protection including racial or ethnic origin, political opinions, political beliefs, religious beliefs, and sexual orientation.

Contact us

To learn more about Information Protection and Business Resilience, contact one of the following KPMG professionals:

Greg Bell
National Practice Leader
Information Protection and Business Resilience
KPMG LLP
404-222-7197
rgregbell@kpmg.com

Doron Rotman
Advisory, Managing Director
KPMG LLP
408-367-7607
drotman@kpmg.com

Michael VanDenBerg
Director, Advisory
KPMG LLP
214-840-8022
mrvandenberg@kpmg.com

kpmg.com

NDPPS 286878