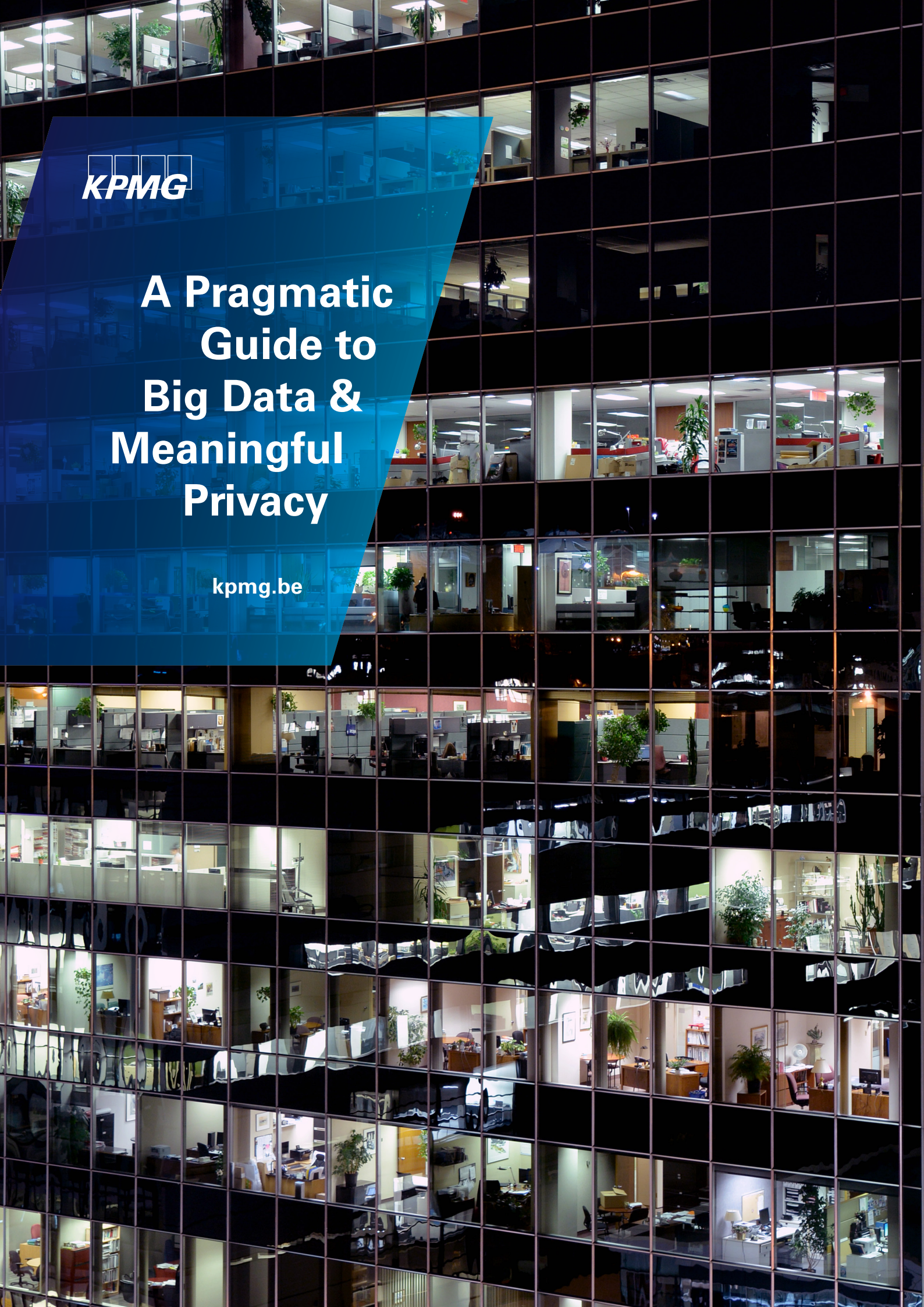


KPMG

A Pragmatic Guide to Big Data & Meaningful Privacy

kpmg.be

From predicting criminal behavior to medical breakthroughs, from location-based restaurant recommendations to customer churn predictions, the benefits of Big Data in everyday life are increasingly self-evident. Organizations similarly see advantages in implementing Big Data programs as part of a strategic business model. There is opportunity to gain a competitive advantage, know their customer better and identify new business areas.

However, concerns about Big Data are also growing. Customers are increasingly wary of the use and potential abuse, of personal information. Regulators are passing and enforcing laws that conflict with its widespread use.

How can your organization balance these demands?

The characteristics of **Big Data**

Companies have collected data for years, and the volume has increased with the technical capacity. In our age of inexpensive processing equipment, the size of the datasets involved are difficult to comprehend. For example, it was reported in 2012, that 90% of all available data was created in the previous two years.[1] To better understand this constantly expanding range, Big Data is commonly defined by three characteristics:

1. *Volume:* data size
2. *Velocity:* speed of change
3. *Variety:* different forms of data sources

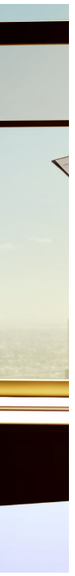
Additionally, Big Data understanding may be supplemented by two additional concepts:

1. *Veracity:* referring to the truthfulness of data
2. *Value:* the business case for using the data.

Due to these characteristics, Big Data sets are too large and rapidly changing to be analyzed using traditional database techniques or commonly used software tools.

[1] "Big Data : turning conceptual thinking into powerful results." KPMG NL (2013).

Big Data is transforming **major industries**



Big Data is not confined to a single industry. KPMG in Belgium (KPMG) clients across all industries recognize Big Data's potential value and seek to harness its capabilities.

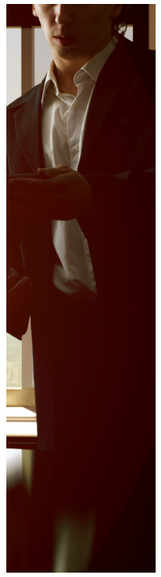
- **Utility companies**, for example, have implemented Smart Grid and Smart Metering technology throughout Europe and globally, resulting in improvements in Outage Response, Renewable Reliability and Load Forecasting , as well as improved maintenance and planning.
- **Retail organizations** use Big Data to improve not only supply chains, but services and products through obtaining a better understanding of their customer behavior and preferences via customer analytics and segmentation.
- **Governments** are starting to use Big Data and analytics to gain insight into citizen matters and requirements, detect fraud or abuse of services and better allocate government resources.
- **Financial institutions** are using Big Data to better identify their customer's markets and assess risk, including creditworthiness, default rates and fraud detection, as well as increasing new product development, and understanding and improving channel usage.

As Big Data continues to transform industries, organizations face a growing demand for talent to help them take advantage of Big Data's opportunities. As a recent KPMG study demonstrates, "organizations which benefit from Big Data & Analytics will have a competitive edge when it comes to better and faster business decisions."

Big Data: **Big risks**

Big Data carries significant informational and business risks that only grow with the data collected. KPMG believes that organizations must find an appropriate balance to opportunities and challenges as they build their Big Data governance programs to enhance Big Data's benefits. We recognize that adopting Big Data requires addressing a number of challenges, including developing a Big Data governance strategy and program to support implementation in manageable steps, while maintaining the larger innovative vision. Meeting these challenges requires integrating the necessary competences, platforms, partnerships and informational changes holistically throughout the organization. Our approach builds compliance issues and social concerns into this ongoing process.

The following sections will focus specifically on the Data Protection risks associated with Big Data programs.



Regulatory compliance and information governance

While Big Data concerns stem from familiar complexities, they also face new issues. A prime challenge with Big Data is compliance with existing laws when the data complexity has increased multifold.

Regulatory agencies have demonstrated concern about the consequences of Big Data accountable as "the risk of consumer injury increases as the volume and sensitivity of data grows."[2] They will hold companies responsible and accountable for non-compliance, therefore companies face the challenge of correctly interpreting current regulations and anticipating future regulations.

The European regulatory privacy landscape is evolving as the European Commission is in the process of implementing Data Protection reform to replace the existing EU Data Protection Directive. Among other requirements, the Data Protection Directive and Regulation demand:

- A "**Right to erasure**" requires personal information previously provided to be removed, including use of this data by third parties. This is a challenge in the Big Data context because multiple forms and copies of data exist, and anonymization is often incomplete.
- **Data minimization**, as a general principle, is in conflict with Big Data's emphasis on processing and analyzing all possible data.
- **Fair processing**, including transparency in data collection, is difficult when data sets are combined from numerous sources and change rapidly.
- **Maintaining the original purpose (intent) of the data**, which may significantly reduce the value of Big Data analytics to reuse data for new and innovative secondary purposes.
- **Safeguarding the data** is always a challenge but is made more difficult when the data is in the form of a valuable and centralized target.[3]

Additionally, the European Union is increasing the consequences of non-compliance with a Data Protection law. The proposed Regulation may include fines of up to 5% of corporate global turnover. The scope of the law will include all companies processing the personal data of EU residents, not only companies operating in Europe.

Reputational, ethical and social challenges

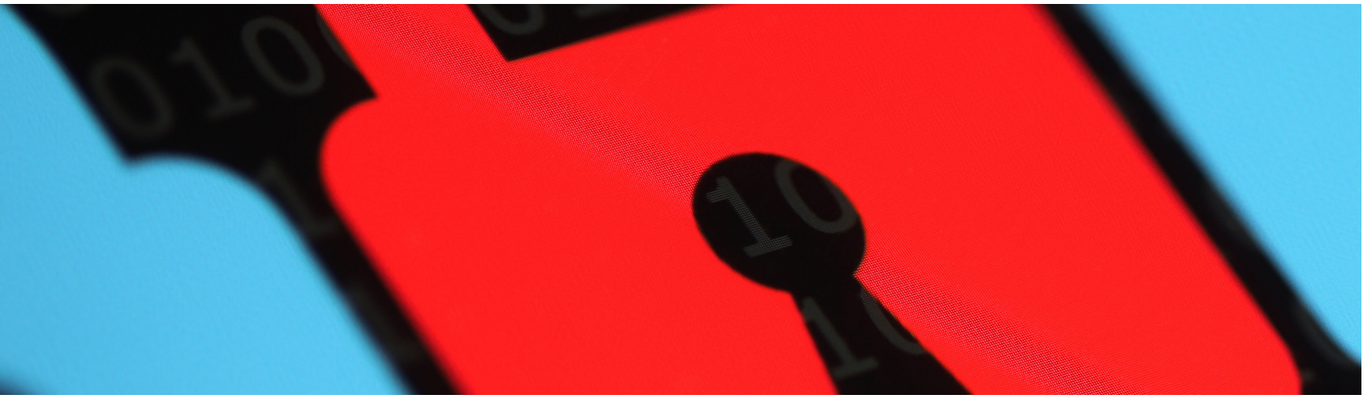
The volume and sensitivity of Big Data sets also creates a need to consider the impact of the analysis conducted. As it becomes integrated into daily life, this Big Data has greater impact on customers and other data subjects and increases awareness of the inherent benefits and dangers. This leads to a reputational risk for companies undertaking Big Data analysis.

The more Big Data is relied upon, the more significant is the resulting analysis. This benefit carries ethical and social weight, as there is an obligation to ensure that the data and the decisions based on it are accurate and fair. It is the responsibility of companies to be socially and ethically aware when using Big Data, or face the reputational risk. Organizations that attempt to implement Big Data initiatives without a strong governance regime in place risk placing themselves in ethical dilemmas without set processes and guidelines to follow.

Some challenges include:

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

[3] UK Information Commissioners Officer, "Big Data and Data Protection" (28 July, 2014)



- **Data-based decision-making** – As data collection and processing ability grow, more decisions are made automatically, without human input. Organizations must decide for which decisions this is appropriate.
- **Behavioral targeting** – The use of information about an individual's past behavior (e.g., web-based searches) or combining offline and online behavior, to select advertisements to display can bring many benefits. If non-transparent it can be invasive or context inappropriate.
- **Data breaches and leaks** – The potential damage of a data breach is multiplied with Big Data, as greater data stores are affected, including data created about a person.
- **Data Duplication** – Duplication is essential to efficient big data processing, however it challenges total information modification or deletion, such as the right to be forgotten.
- **Dataset Combination** – Increased value can be obtained from data by combining it with other data sets. Yet such a combination may reveal unwanted or intrusive information about data subjects.
- **Re-identification risk** – Fully anonymized data removes the risks associated with personal data. However, incomplete anonymization may result in compromised privacy as the data is combined with previously collected, complex data sets including geo-location, image recognition and behavioral tracking.
- **Sharing Data with Third Parties –** Data sharing may greatly increase the data value for all parties. However, the security and privacy data protections in place at all third-parties must be adequate to protect the combined data.

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

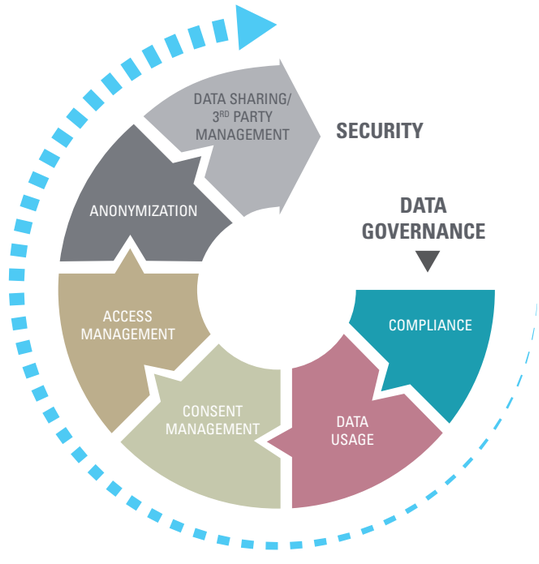
KPMG has a two-pronged approach to building a Big Data Privacy and Security Program. We believe that both the technical and governance layers are essential to success.

We have identified **eight high-risk focus areas** and offer guidance based on an augmented Information Governance life cycle and a Big Data platform to assist organizations in minimizing risk and maximizing control over Big Data.

A strong Big Data Privacy and Security program requires two complementary approaches. First there must be strong Governance Processes in place to determine data usage. Second, these processes must rely on a Big Data Platform that is capable of maintaining the privacy and security requirements of the program. The eight areas highlighted below align with one or both of Governance or Platform.

1. *Data governance and retention – Governance*

   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 which includes a strong ethical code, along with process, training, people, and metrics is imperative to govern what organizations can do within a Big Data program.



   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 – Governance*

   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 – Governance*

   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 – Governance & Platform*

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 it 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, not only at a larger account level.

5 *Access management – Governance & Platform*

The amplified size and scope of Big Data increases the significance of granting access to it. Organizations must control who has access to the data sets both internal and external to the organization. This requires a comprehensive access management regime of users including third parties and partners. All new user access requests should require approval, and access must be removed immediately when it is no longer required. Organizations should perform periodic reviews of internal and external existing user access, and adopt segregation of duties where access to systems is based on job function. Organizations can automate this 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 *Data sharing/third-party management – Platform & Governance*

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 limit their Big Data partner's access, and restrict combinations of data sets that are considered too sensitive. Organizations must include Big Data with the overall evaluation of third-party performance to ensure ongoing monitoring of compliance with data-sharing agreements.

7. *Anonymization – Platform*

Anonymization means processing data so that it is irreversibly de-identified, through the removal of identifying information. Data anonymization is critical to maintaining the privacy of the original data set when considering the long-term use of Big Data. Achieving true 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.

8. *Security practices – Governance & Platform*

Organizations should recognize the potential value of their combined data sets and work to make them less attractive targets to unauthorized parties i.e. hackers. This can be done by increasing security safeguards, storing information in such a way that its usefulness is minimized or outsourcing certain security risks to third parties.

How KPMG can help

Based on our assessment that Big Data programs rely both on strong Governance Process and appropriate Technical Platforms, KPMG finds that clients benefit from support aligned with these perspectives:

- **Information governance processes:** Big Data program development is predicated upon a strong sense of client requirement for their program and how to best implement it while complying with privacy and security requirements. KPMG supports clients in creating and implementing the appropriate policies and process while advising on third party management.
- **Technical platform:** KPMG advises our clients on the appropriate technical solutions for their needs, including various forms of anonymization and pseudo-anonymization, data models, third party partners, etc.

Understanding our client's need to minimize the attractiveness of their Big Data sets to outsiders and outsource risk, KPMG has also developed a tailored privacy-enhancing processing platform, KAVE. For more information about this open source platform, please contact us.

Case Study:
Boosting Big Data with benchmarking

Client:
Global Telecom Company

Challenge:
Client sought to implement industry best 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 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:
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 insights to your organization.

A sampling of our previous engagements includes:

- **Big Data Analytics Enhancement:**
  Provide additional capabilities and resources to an analytics platform to improve your Big Data insight development and program development.
- **Big Data Security and Privacy Program Assessments:**
  Define and document 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 future state road map to enable an 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 the current Big Data analytics capabilities and identify opportunities for development and improvement

Glossary

Anonymization
process to remove all personally identifiable information from a data set and permanently turn it into non-identifiable data.

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.

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

Personal Data
information about a person. It may uniquely identify a person or may be associat0ith many people.

Pseudo-Anonymization
data that is purposely not fully anonymized. Individuals cannot be identified from it in its current state but with additional information such as a key, re-identification is possible. A recommended approach to reducing the risk associated with data collection and analysis.

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 & beliefs, religious beliefs and sexual orientation.

Contact us

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

Bart Meyer
Partner, KPMG Advisory
T +32 3 821 1780
E bmeyer1@kpmg.com

Benny Bogaerts
Director, KPMG Advisory
T +32 3 821 1893
E bbogaerts@kpmg.com

Peter van den Spiegel
Senior Manager, KPMG Advisory
T +32 2 708 3779
E pvandenspiegel@kpmg.com

Sarah Pipes
Data Protection Specialist, KPMG Advisory
T +32 3 821 17 07
E sarahpipes@kpmg.com