Big Data Analytics & Privacy: How To Resolve This Paradox?

Ronald Koorn, Jeffrey Bholasing, Sarah Pipes, Doron Rotman, Christopher Kypreos, Souella Cumming, Ashleigh van Kerckhoven, Koji Hijikata and Takayuki Manchu

Mind the hype. Big Data is the next big thing in the world of IT, partly as people are becoming increasingly aware of its potential to impact the business world and society as a whole. In addition, people are awaking to the powerful possibilities that combining structured and unstructured, internal and external data holds. The opportunity this offers to perform advanced analytics, or "Big Data Analytics", provides insight into the various outcomes of Big Data Analytics, and particularly the impact it could have on the Privacy of individuals. In this article, we elaborate on the Privacy challenges with Big Data by looking at Big Data applications from four continents, and examine how the respective legal frameworks relate to those Big Data uses.

R.F. Koorn
is a partner at KPMG Netherlands.
koorn.ronald@kpmg.nl

J.R. Bholasing
is a manager at KPMG Switzerland.
jeffreybholasing@kpmg.com

S. Pipes
is a manager at KPMG US seconded to KPMG Belgium.
spipes@kpmg.com

D. Rotman
is a managing director at KPMG US.
drotman@kpmg.com

C. Kypreos
is a senior associate in the KPMG Cyber practice.
ckypreos@kpmg.com

S. Cumming
is an Advisory partner and head of Government Services for KPMG New Zealand.
smcumming@kpmg.co.nz

A. van Kerckhoven
is a manager at KPMG South Africa.
ashleigh.vankerckhoven@kpmg.co.za

K. Hijikata
is a director at KPMG Japan.
koji.hijikata@jp.kpmg.com

T. Manchu
is a senior manager in the Cyber Security practice at KPMG Japan.
takayuki.manchu@jp.kpmg.com

Introduction

From predicting criminal behavior to gene-based medical breakthroughs, from location-based restaurant recommendations to customer churn predictions, the benefits of Big Data in everyday life are becoming self-evident.[1] Likewise, organizations see advantages in applying so-called "datafication"[2] and implementing Big Data programs as part of strategic business models. There are opportunities to gain a competitive advantage, to get to know your customer behavior better and to identify new business areas.

However, concerns about Big Data are also growing. Customers are increasingly wary of the unlimited data hunger, extensive use, and potential abuse, of personal information by profiling and being provided personalized online ads or even personalized prices based on online tracking. Regulators are enacting and trying to enforce laws that conflict with widespread personal data use.

In this article, we address successful and unsuccessful applications of Big Data from a Privacy perspective. After describing some examples of how organizations are currently using Big Data, we look at the Privacy challenges with Big Data in different geographic areas. First, the legal Privacy framework in each region is discussed, followed by an example of how a Big Data implementation led to a Privacy failure. We conclude by presenting a number of potential solutions and conclusions with a view to dealing with Big Data in a Privacy-compliant manner.

1 Big Data is high-volume, high-velocity & high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight & decision-making (Gartner).

2 Datafication is a modern technological trend turning many aspects of our life into computerized data and transforming this information into new forms of value (see also [Maye13]).

How Are Organizations Currently Utilizing Big Data?

Companies have been collecting data for years, and as the technical capacity increased, so did the volume. 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 had been created in the previous two years ([John12]).

Big Data sets are either too large or too versatile to be analyzed using traditional relational and multidimensional database techniques or commonly used software tools to capture, manage, and process the data within a reasonable elapsed time ([ISAC13]).

Big Data is not confined to a single industry. Organizations across all industries recognize Big Data's value and seek to harness its potential. For example (see also [Klou15] and [Maye13]):

- *Retail companies* use Big Data to improve services and products through customer analytics, customer behavior, sentiment analysis, profiling and segmentation.
- *Governments* use Big Data analytics to gain insight into citizen activities and requirements, detect fraud or abuse of services, focus "nudging" on specific behavior, strengthen national security and better allocate government resources.
- *Financial institutions* are using Big Data to better identify their customer markets and assess risks, including creditworthiness and default rates.
- The *travel industry & public transport sector* is identifying current and potential loyal and high value customers, by analyzing travel patterns and frequency across both personal and professional platforms. Recent analytics predict travel delays before they even happen.
- *Insurance companies* are introducing risk-based ratings (with higher premiums) and efficiencies through the use of Big Data analytics, even by placing in-vehicle data recorders for analyzing driving behavior and events in real time. Based on such analysis and profiling, most claims may be processed automatically through Straight-Through Processing, whereas others are set aside for additional manual review.

As Big Data continues to transform industries, organizations face a growing demand for talent to help them take advantage of Big Data's opportunities. 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" ([KPMG13]). However, most organizations are still struggling to make use of Big Data effectively, due to a lack of high quality data, skilled data scientists, tooling, integration, and well-founded decision-making ([KPMG14]). Currently, most Big Data analytics are focused on operational and tactical questions; in effect, not many strategic or social issues have been addressed by private or public organizations.

Example of successful Big Data – Europe

An example of a successful Big Data implementation concerns the coronation of the new King in the Netherlands on 30 April 2013, when Amsterdam experienced one of its busiest days in history. Throughout the city, multiple performances and other events were organized, attracting an estimated total of 750,000 visitors to the inner city of Amsterdam ([ANP13]).

Holland's largest news website placed a widget online that visualized the intensity and movement of the crowds in the streets. Important to the concept of the widget was that of "Privacy by design". The widget functioned by receiving antenna data from the mobile telecommunication network on the number of users connected to a cell.

The user's Privacy was safeguarded as only aggregated information from the antennas was gathered that merely specified the number of connections for each antenna at each point in time. The data did not specify any additional information about the users of the telecom network and therefore the data could not be related to an identifiable individual.

Figure 1. **Crowd control visualization based on Data Analytics.**

What Are the Privacy Challenges With Big Data?

An argument that soon gained impetus as the attention for the potential of Big Data was increasing, was the impact it could have on the Privacy and Security of individuals. The massive collection of data and the interrelations that can be derived with Big Data, could lead to detailed insights into locations, specific consumer interests and human behavior for Big Data processors, something the individuals concerned might not even be aware of but would be perceived by them as a breach of their Privacy.

When it comes to using data analytics effectively for predicting human behavior, there are quite a few challenges ([Str015]):

- Human psychology of cognitive inertia: people are inclined to resist change and to revise their assumptions.
- Information overload and quality: more data does not mean better data, especially when internal and external data sources are combined. Analyzing unbalanced data or data without the context may lead to prejudices, bias or incorrect conclusions.
- Ability to make sense of behavioral data: deriving value from large volumes of data, for instance when analyzing data about emotions and sentiments, which are "scraped" off social media by market research companies.

As a consequence, even while data scientists are attempting to outsmart us, individual human behavior might be too complicated to predict accurately.

Privacy Legislation Affecting Big Data

The processing of personally identifiable information (PII)[3] is subject to strict regulations in most, if not all parts of the developed world, in order to protect the Privacy of individuals. Making use of Big Data is a means of processing and is therefore subject to Privacy regulation in cases where PII is concerned ([Vier12]). For instance, in the EU the current Data Protection Directive requires organizations that use automated decision-making to "produce legal effects", such as not granting a mortgage based on programmed rules, so as to provide notice and to allow for human intervention.

In some parts of the world, applicable laws have been updated with a view to responding to collection practices made possible by new technology, namely data-gathering tools such as social media and mobile applications.[4] The draft EU General Data Protection Regulation specifically mentions profiling, and that explicit consent is needed when profiling can significantly affect individuals (by positive as well as false negatives). If the data is anonymized or pseudonymized or profiling serves the "performance of a contract" or adequately safeguards the "legitimate interests" of individuals, in which case explicit consent is not required by law. Obtaining and maintaining explicit consent from each individual is usually an administrative hassle and – as most online users automatically click the Agree button to give their permission – is not as strong as an organization proving its overall Privacy accountability ([EDPS15]).

Under most Privacy regimes, the processing of PII is only allowed in circumstances where there are unambiguous legal processing grounds. Important in the processing of PII is that the cause and processing ground need to be established up front, in order to process PII in a compliant manner. One or more legal grounds need to be established per processing objective. In such a case, the ground for processing only concerns the processing for that objective. For example: data which is gathered with a view to carrying out a mutual agreement, such as billing by a telecom provider, may not be used for an additional, not closely related cause, like using Big Data methods to determine which (types of) callers call most frequently depending on varying marketing triggers. In short, PII data that is originally gathered for a legitimate purpose may not be reused for Big Data purposes without a separate processing ground or a purpose that is sufficiently "comparable" or "compatible". In practice, this often poses a problem for the party aiming to use Big Data techniques ([Moer14]).

3 The original scope of 'Personally Identifiable Information' was sufficiently broad when the European Privacy regulation was introduced: in certain circumstances it would include IP addresses ([Zwen13]). Since then, Working Party 29 have published a number of opinions related to the scope of the definition 'personally identifiable information' (http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf). It can be concluded that the scope of the definition of personally (identifiable) information as used within the European Commission, is becoming increasingly wider ([Cuij12]).

4 The FTC revised the COPPA requirements in 2013 to include Third Parties such as advertising networks covering 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."

Obtaining online customer consent for data analytics has degenerated to a clickable Privacy control

A solution that is often sought by parties willing to use Big Data in order to overcome the restrictions of Privacy legislation, is anonymization (see also section "Potential Solutions to Big Data Privacy Challenges"). Since Privacy legislation is only applicable to PII, simply removing or "masking" the PII fields such as name, address, birthdate and any ID numbers, may appear a good solution for avoiding the restrictions of Privacy legislation. However, in practice this solution will often prove inadequate ([EC14a]). First of all, with a large dataset, spontaneous recognition of an individual may still be possible ([Berk14]). Furthermore, not only directly identifiable data, but also indirectly identifiable data ("indirect inheritance"), fall under the scope of Privacy regulation. Data elements such as name and address are considered to be direct PII. Data elements that cannot be linked directly to an individual but can be traced back to an individual by performing deeper analyses on the data or by linking with additional data elements, are considered indirect PII. This indirect PII is also subject to all requirements in current and upcoming Privacy legislation.

Potential Solutions to Big Data Privacy Challenges

Based on the legal Privacy framework in the light of Big Data and the examples in different parts of the world, we can identify a number of potential controls for dealing with Big Data in a Privacy-compliant manner. These are:

Data Governance & Retention
Underlying any Big Data program, a data governance structure must be established that provides clear directions at to how the data is to be handled and protected by the organization. This program begins with a clear organizational structure around data governance (e.g., who is data owner or data steward/custodian, who is responsible for protecting the data), followed by additional key components such as policies, standards and procedures, including critical elements such as data use monitoring, data retention and data destruction. Preferably, data governance is translated and codified in a standard, in guidelines or in a framework (de-facto or de-jure). Several organizations are working on Big Data Privacy standards and guidance, such as ISO/IEC (see [ISO14]), NIST/IEEE (Big Data Public Working Group), Cloud Security Alliance (Big Data Working Group), UK's Information Commissioners Office ([ICO14]), Working Group 29 of the European Commission ([EC14b]), and the Information Accountability Foundation ([IAF14]). Another avenue is the development of a Professional Code of Conduct for data scientists to stimulate ethical data use and to avoid data analytics cherry-picking.

Compliance
Organizations must identify and understand the Privacy and – in some countries – (cyber) security 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 overview when it comes to the monitoring of compliance with regulatory and contractual commitments.

Compliance requires the development of a comprehensive Data (Privacy) Control Framework and risk-based roadmap

Big Data & Privacy Case Study US – Target

Target is a well-known and early user of Big Data's predictive powers, having correctly predicted a teenager's pregnancy before her parents were even aware of the addition to the family ([Hill12]). Target's datasets were therefore a highly visible and valuable source of information, and the subsequent breach it suffered of over 70 million customer records (including 40 million credit card numbers) resulted in significant damage to its finances and reputation. Target's systems were breached by hackers using credentials stolen from a third party providing heating, ventilating, and air conditioning (HVAC) services, and the presumed Russian hacker has not yet been identified. Target's security team received an alert that there was malware on their systems; however, they did not act on the notification and soon millions of records were leaking from their systems. Senator John Rockefeller stated "I think we can all agree that if Target – or any other company – is going to collect detailed information about its customers, they need to do everything possible to protect it from identity thieves" ([Rise14]). The CEO was fired; its shoppers were slow to return; and its data-breach costs have climbed to a quarter-billion dollars ([SCTi14]).

Currently, the US White House has issued a report on Big Data and Privacy ([Whit14]) and is proposing a federal Privacy Bill of Rights.

for implementation. Companies can then take advantage of automated controls based on such a standard and transition from manual efforts to ensure ongoing compliance.

In a previous article, we discussed the different Privacy by Design options that can be applied for automating (preventative) Privacy controls ([Koor11]).

Data Use Cases & 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 such as a Privacy Impact Assessment prior to the adoption of new data feeds. Key considerations in the adoption of any new data feed are the quality of the dataset and the potential risk for re-identification, which increases when existing data feeds are combined with other data feeds. With respect to the limited quality of datasets involving human behavior, an upcoming approach is to just add substantially more data in order for this large quantity to resolve the quality problem. This approach is practiced and studied by researchers at Google ([Hale09]).

Consent / Permission Management
Customer consent management is critical to a successful implementation of any Big Data governance regime. Consent is not required for every Big Data project, so if the organization can prove "compatible use" or "legitimate interest" (see [EC14b]), it may do without. Unfortunately, obtaining online customer consent has degenerated to a clickable Privacy control.

Big Data & Privacy Case Study EU – TomTom

A Big Data project with geo-location data of individuals has led to serious Privacy challenges under the EU Privacy regime. De Montoye et al. demonstrated that when dealing with individuals' geographic location data, in a dataset where the location of an individual is specified hourly, the vast majority of the user's identity can be derived based on less than five different location measurements ([Mont13]). This clearly shows how data which may not be perceived as such, could in fact be indirect PII and thus in scope for the EU Privacy legislative framework.

Dutch navigation provider TomTom experienced a Privacy challenge with the processing of geo-location data first hand, when Dutch DPA ruled that TomTom was in breach with EU Privacy legislation. TomTom subsequently received public criticism for not paying due respect to users' Privacy. For the purpose of navigation, the TomTom device obviously processes the user's geo-location data. Apart from the processing during navigation, historical data of previous use is also stored in the TomTom device. You could argue that if all data is retained in the device, there can be no major Privacy risk. Yet, newer TomTom devices are connected via the Internet and upload the user's geo-location real-time to a central server in order to calculate traffic intensity and provide better navigation advice. With previous TomTom devices that are not online, the user's historical geo-location data is uploaded when the user connects the device to the computer to perform updates.

The TomTom devices asked for the user's consent to process geo-location data, but the Dutch DPA ruled that the consent did not suffice, as it was not specific enough. In consequence, there was no legitimate ground for the processing and TomTom was in breach of legislation. TomTom followed up immediately by requesting new consents to the whole user community where this applied. The Dutch DPA subsequently ruled that TomTom now possessed the correct consent for processing the relevant data.

TomTom experienced another Privacy incident when it was publicized that it was selling traffic data including historical speed to several third parties. Data recipients included local and national governments. It emerged that the Dutch police also used that dataset as an additional argument for placing speed cameras at locations where the average speed was too high. However, an investigation by the Dutch DPA showed that all traffic data was anonymized and aggregated, and that no individual data was sold. Even so, TomTom decided to alter their licensing agreements with data recipients to curtail this type of usage. Currently, TomTom is a leader with its privacy practices, according to the Mobility Data Analytics study by the Government Accountability Office ([GAO13]).

Customer consent requires the following components:

- Transparency – Organizations should provide their customers with a clear understanding of the information the organization collects and how it will be used.
- Consistency – Organizations should provide consistent consent mechanisms across all products, and register Big Data preferences upfront.
- Granularity – Organizations should allow customers to provide or withdraw their permissions at the individual device level, not only at a larger account level.

Access Management

Given the amplified size and scope of Big Data, organizations must effectively control who in the organization has access to the datasets. This requires a comprehensive and granular access management regime including review and approval of new user access requests and periodic reviews of existing user access to ensure that privilege requirements meet security and compliance requirements. Finally, organizations should adopt segregation of duties where access to systems is based on job functions. Organizations can automate the process by using policy engines or access management tools to implement Role- or Attribute-Based Access Controls (RBAC resp. ABAC). This will help them make dynamic access decisions and integrate with existing tools and directories for provisioning and certification. Each extensive data warehouse with identifiable personal data is a Privacy risk in itself ("single point of success or failure"), and without strong access controls and segmentation, a single disgruntled employee or hacker might create a major Privacy breach.

Anonymization & Pseudonymization

Anonymization means removing all personally identifiable information (PII) from a dataset and permanently

Big Data & Privacy Case Study New Zealand – Sensing City

Data protection in New Zealand is governed by the Privacy Act 1993, which exists alongside a range of industry codes including health, telecommunications and credit reporting, and other relevant legislation such as the Statistics Act 1975. The New Zealand Privacy legislation is principle-based, being influenced by the 1980 OECD Guidelines on the Protection of Privacy and Trans-border Flows of Personal Data. The European Union recognized the New Zealand Privacy Act as providing an adequate standard of data protection for the purposes of European law ([EC12]).

However, the advent of Big Data presents some challenges to these Privacy principles, especially as it is becoming easier to re-identify personal information from anonymized datasets and predictive modelling. A 2011 review by the New Zealand Law Commission did not recommend any fundamental changes to New Zealand's principle-based legislation, but has recommended a number of changes which, in the words of the New Zealand Privacy Commissioner, "would power up Privacy law to meet the challenge of protecting New Zealanders' personal information in the digital age." ([Priv11])

In December 2013, the Ministers of Finance and Statistics established a New Zealand Data Futures Forum to explore the potential benefits and risks for New Zealand of sharing, linking and using data. Privacy is recognized as an integral part of the future use of Big Data. In its July 2014 report, the New Zealand Data Futures Forum concluded that "there are huge economic, social, environmental and personal opportunities for New Zealand if we can strengthen trust, control and inclusion in the data-use ecosystem."

How Big Data can be used in a Privacy-compliant manner became apparent when the "Sensing City" was designed. Not only did the devastating earthquakes that hit the Christchurch and Canterbury region of New Zealand in 2010/2011 destroy the lives and livelihoods of tens of thousands, they also caused widespread damage to infrastructure and property in the region. Innovation and future-proofing the city have become some of the underlying design principles of the extensive rebuilding program. Planners have looked at the work of architect Roger Dennis to develop Christchurch as a Sensing City. The Sensing City initiative is based on transforming raw (Internet of Things) data centered on people and objects within the city into valuable information products. Data is combined and analyzed to drive the future applications, services and decisions in a wide range of sectors. Real-time open data provided by Sensing City has the potential to add immense value to local business and the lives of Christchurch citizens through the range of services it creates.

turning it into non-identifiable data. Well-designed anonymization or data-masking is critical to the long-term use of Big Data while maintaining the Privacy of the original datasets. In practice, achieving full anonymization in large datasets is very challenging; the outcries about the publicly accessible and anonymized internet search data (AOL in US) and health data (care.data of the National Health Service in the UK) demonstrate that by means of data mining and cross-referencing several identities could be derived. Therefore, using strong anonymization close to the data source and quantifying the risk of re-identification should be part of each organization's approach ([Koot12]). Alternatively, the entire data analytics can in some instances be positioned at the data source, which means the results exchanged for further central processing are identity-free.

Big Data & Privacy Status – Africa / South Africa

Unlike the EU, there is no regional mandate within Africa for countries to implement laws ensuring the protection of personal information. Most African countries have yet to introduce such Privacy legislation. However, 17 countries that have introduced Privacy legislation have adapted legislation that is by and large similar to the Privacy laws found in Europe, Latin America and Asia. These laws are still in their foundational stages, due to the lack of established regulators. Furthermore, the lack of regional guidance for the development of Privacy laws has resulted in some inconsistencies between Privacy laws in Africa in respect of the legal basis for the collection and use, database registration, breach notification, and the cross-border transfer of personal information ([Rich14]).

The right of individuals to Privacy is founded in the Constitution of the Republic of South Africa Act 108 of 1996; it is supported by common law and various other pieces of legislation. However, although comprehensive Privacy legislation has been laid down in the Protection of Personal Information Act 4 of 2013 (POPI Act), it is yet to become effective in full (Note: some sections of the POPI Act relating to the creation of the Information Regulator and the guidelines for the development of data protection regulations are now effective). The POPI Act prescribes the conditions for the lawful processing of personal information by both public and private entities. Although the legislation has in effect been passed, there is no obligation for organizations to comply. Once the POPI Act is effective in its entirety, organizations will have a one-year grace period to comply with its conditions.

Although the POPI Act is rather similar to European and UK Privacy laws, being based on the OECD guidelines for the fair processing of personal information in respect of: accountability, notice, choice and consent, purpose specification, security safeguards, data quality and participation, it has some unique characteristics that play an important role in Big Data Analytics, including the following:

- requirement for the protection of personal information that uniquely identifies both natural and legal persons;
- requirement that only the minimal amount of personal information be collected to achieve a specific business purpose;
- specific circumstances under which specific personal information (such as health information, sexual orientation, political affiliation, and race) , including children's personal information, may be processed;
- minimum requirements for the transfer of personal information across borders;
- requirement for fair treatment of individuals in automated decision-making processes (e.g., deriving a risk profile for an individual through automated processes);
- obligations for organizations to notify and obtain authorization from the Information Regulator for certain processing activities.

Within the South African context, organizations are yet to realize the benefits of Big Data, as there are only few organizations using Big Data for strategic decisions, risk management, and/or predictive modelling. Currently, South Africa remains relatively immature when it comes to the utilization of Big Data by organizations and its Privacy framework. The legal framework provides opportunities for Big Data use, however. This, in its turn, presents a unique opportunity for South African organizations to begin building their Big Data capability in conjunction with the upcoming Privacy legislation and use the conditions for the lawful processing of personal information, as prescribed in the POPI Act, in the context of their Big Data initiatives.

Halfway fully anonymized data and identity-rich personal data is the concept of pseudonymization. Pseudonymization refers to a procedure in which identifying data will be replaced with a specific, non-reversible hashing algorithm by encrypted data (pseudo-ID). This algorithm can always calculate the same pseudo-ID for the same person, so that information about that person, derived from different sources or over a longer timespan, can be combined (longitudinal research). This differentiates pseudonymization from anonymization, where combining personal data is impossible.[5]

5 Based on ISO standard TS 25237.

While the concepts of anonymization and pseudonymization is complex and few companies disclose 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 datasets. A trusted third party can be an organizational solution for pseudonymization, thereby de-coupling the identity data from the other personal data.

Big Data & Privacy Status – Japan

Big Data as well as regulatory Privacy developments in Japan are still in their early stages and appear slow in comparison with other countries. To counteract this, the Japanese ministry of Economy, Trade and Industry (METI), which is in charge of Big Data utilization, has developed a national strategy to accelerate utilization and increase international competitiveness driven by Big Data. On June 2014, METI established the "Strategic Council for Creating Data-Driven Innovation", in which more than 200 enterprises participate, to accelerate the collaboration between the government and enterprises with regard to data utilization.

As a result of the increased efforts of METI and Japanese enterprises, a number of Big Data initiatives have recently been developed. Some of these initiatives amount to the following:

- analysis of client-activity logs for an automatic selection of advertisements based on client attributes – by a web service company,
- automatic abnormal pattern-matching in production line for a cost reduction of defects – by a precision equipment manufacturer,
- vehicle traffic information analysis for the provision of accurate traffic information in case of incidents – by an automotive manufacturer.

One of the reasons for the slow uptake of Big Data in Japan is that enterprises are reluctant to use personal data due to the negative popular view. Another reason is that the Japanese legislation regarding Privacy-compliant use of Big Data is ambiguous and still subject to change.

The Japanese Act on the Protection of Personal Information was developed in 2003 and became effective for the private sector in 2005. The act stipulates the responsibilities of national/local governments and obligations of data processors when it comes to protecting personal information, but not the Privacy rights of data subjects. The Act is supplemented by guidelines established by various ministries in charge of their respective industries. In the Act, no rules for the international transfer of personal data are included.

Since January 2015, the congress has been in discussion to modify the Act to ensure that it covers how personal data should be used by various businesses. In addition, alignments with major Privacy and data protection regulators in other countries are currently taking place. Congress is proposing to create a definition of non-personal data and an organization to interpret the fine line between personal data and non-personal data on a case-by-case basis and to regulate obligations for handling non-personal information, which would be less restrictive than personal information.

In addition to the Act on the Protection of Personal Information, Japan has the Act on the Use of Numbers to Identify a Specific Individual in the Administrative Procedure. The objective of this Act is to make administrative procedures efficient by using an Individual Number ("My Number" in Japanese) and a Corporate Number for identification. As of October 2015, an Individual Number is assigned to individuals who have Japanese residency. As the Individual Number is categorized as Specific Personal Information, it will also be subject to the Act on the Protection of Personal Information.

Data Sharing / Third-Party Management

Big Data concerns grow exponentially as the data is combined with additional datasets. Organizations maintain a responsibility toward 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 their overall strategy for the performance of third-party assessment and audits to ensure ongoing monitoring of third parties with a view to complying with data-sharing agreements.

Conclusion

Data abundance will become epidemic with all above-mentioned developments, and the billions of (Internet of Things) sensors will also bombard us with additional data, including personal data. The opportunities that (Big) Data Analytics present are too good to miss out on. Privacy is often an argument that is mentioned in a Big Data context, but for many it is unclear what Privacy regulations stipulate, how they relate to Big Data and what measures could be taken in order to use Big Data without creating Privacy risks or even Privacy breaches. As we have outlined, the different legal frameworks applicable to Big Data differ per region or country, but are converging in OECD countries. Ultimately, we have presented a universally applicable approach for using Big Data in a Privacy-compliant manner. The specific implementation should be based on a thorough Privacy Impact Assessment of the Big Data project. Big Data is here to stay and so is Privacy; the two can co-exist and even be married up – and thereby resolve the paradox in this article – by addressing Privacy aspects early on in the design phase of Big Data initiatives.

References

[ANP13] ANP, *Meer dan 700.000 bezoekers in Amsterdam*, 1 May 2013, http://www.nu.nl/binnenland/3411540/meer-dan-700000-bezoekers-in-amsterdam.html.

[Berk14] J. Berkvens & J. Prins, *De bescherming van persoonsgegevens*, Recht en Computer, nr. 4, 2014, p. 3.

[Cuij12] C.M.K.C. Cuijpers and P. Marcelis, *Oprekking van het concept persoonsgegevens – beperking van privacybescherming?*, Computerrecht, 2012 (13), pp. 339-351, https://pure.uvt.nl/ws/files/1477196/182_concept_persoonsgegeven_cc_1_.pdf

[EC12] European Commission, Commission Implementing Decision, 19 December 2012.

[EC14a] European Commission, Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymisation Techniques, Brussels: WP29, 2014.

[EC14b] European Commission, Article 29 Data Protection Working Party, Opinion 06/2014 on legitimate interests, 2014, http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf.

[EDPS15] European Data Protection Supervisor, *Opinion 7/2015: Meeting the challenges of big data*, 19 November 2015.

[GAO13] GAO, *In-Car Location-Based Services: Companies Are Taking Steps to Protect Privacy, but Some Risks May Not Be Clear to Consumers*, GAO-14-81: December 2013. http://www.gao.gov/products/GAO-14-81.

[Hale09] A. Halevy, P. Norvig and F. Pereira, *The Unreasonable Effectiveness of Data*, IEEE Intelligent Systems, vol. 24, no. 2, March/April 2009.

[Hill12] K. Hill, *How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did*, Forbes, 16 February 2012, http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did.

[IAF14a] Information Accountability Foundation, *A Unified Ethical Frame for Big Data Analysis*, version 1, 2014, http://information-accountability.org/wp-content/uploads/IAF-Unified-Ethical-Frame-v1-08-October-2014.pdf.

[IAF14b] Information Accountability Foundation, *Big Data Ethics Initiative Part-B – Interrogation Framework*, draft version, 2015, http://informationaccountability.org/wp-content/uploads/IAF-Big-Data-Ethics-Initiative-Draft-Part-B-Final-03-03-2015.pdf.

[ICO14] Information Commissioners Office, *Big data and data protection*, 2014, https://ico.org.uk/media/for-organisations/documents/1541/big-data-and-data-protection.pdf.

[ISAC13] ISACA, *Big Data: Impacts & Benefits*, ISACA White Paper, March 2013.

[ISO14] ISO/IEC JTC 1 Big Data Steering Group, *Big Data: Preliminary Report 2014*, 2014, http://www.iso.org/iso/big_data_report-jtc1.pdf.

[John12] J. Johnson, *Big Data + Big Analytics = Big Opportunity*, Financial Executive, July/August 2012.

[Klou15] S. Klous and N. Wielaard, *Wij zijn Big Data: de toekomst van de informatiesamenleving*, Business Contact, 2015.

[Koor11] R. Koorn & J. ter Hart, *Privacy by Design: from Privacy Policy to Privacy-Enhancing Technologies*, Compact 2011/0 (international edition)

[Koot12] M.R. Koot, Concept of k-anonymity in PhD thesis "Measuring and predicting anonymity", 2012, http://dare.uva.nl/document/2/107610.

[KPMG13] KPMG, *Big Data & Analytics: turning conceptual thinking into powerful results*, KPMG Netherlands, 2013.

[KPMG14] KPMG, *Going beyond the data: Achieving actionable insights with data and analytics*, KPMG International, 2014.

[Maye13] V. Mayer-Schönberger and K. Cukier, *Big Data: A revolution that will transform how we live, work, and think*, Maven Publishing, 2013.

[Moer14] L. Moerel, *Big Data Protection: How to Make the Draft EU Regulation on Data Protection*, Tilburg University, 2014.

[Mont13] Y. de Montjoye, C. Hidalgo, M. Verleysen & V. Blondel, *Unique in the Crowd: The Privacy Bounds*, Scientific Reports, Volume 3, 2013.

[Priv11] Privacy Commisioner, *Media release: Law Commission's Package of Privacy reforms Would Build Trust and Certainty*, August 2011, https://www.Privacy.org.nz/news-and-publications/statements-media-releases/media-release-law-commission-s-package-of-Privacy-reforms-would-build-trust-and-certainty/.

[Rich14] C. Rich, *Privacy Laws in Africa and the Middle East*, Privacy & Security Law report, April 2014.

[Rise14] T. Risen, *FTC Investigates Target Data Breach*, U.S. News & World Report, 26 March 2014, http://www.usnews.com/news/articles/2014/03/26/ftc-investigates-target-data-breach.

[SCTi14] *SC Times, A year after Target data breach, aftershocks finally end*, 25 November 2014, http://www.sctimes.com/story/money/2014/11/25/year-target-data-breach-aftershocks-finally-end/70080462.

[Stro15] C. Strong, *Humanizing Big Data: Marketing at the Meeting of Data, Social Science and Consumer Insight*, Kogan Page Ltd., 2015.

[Vier12] L. Viergever & J. Koëter, *Is onze privacyregelgeving 'Big data proof'?* Tijdschrift voor Internetrecht, nr. 6, 2012.

[Whit14] White House, *Report to the President: Big Data and Privacy: A Technological Perspective*, May 2014, https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_Privacy_-_may_2014.pdf.

[Zwen13] G.J. Zwenne, *Diluted Privacy Law*, inaugural lecture to the office of Professor of Law and the Information Society at the University of Leiden, 12 April 2013, https://zwenneblog.weblog.leidenuniv.nl/files/2013/09/G-J.-Zwenne-Diluted-Privacy-Law-inaugural-lecture-Leiden-12-April-2013-ENG.pdf

About the Authors

R.F. Koorn is a partner at KPMG Netherlands and is responsible for privacy services. His focus areas are Privacy, IT/Data Governance, e-invoicing and the flexibility of IT, especially at public sector and educational organizations and IT service providers. He co-developed the Dutch Privacy Audit Framework and served on ISACA's Global Privacy Task Force. He also spent two years at KPMG US.

J.R. Bholasing is a manager at KPMG Switzerland. He combines his IT and law degrees to bridge the gap between business and support functions (IT, legal, etc.). He conducts part-time PhD research on the subject of Big Data and Privacy.

S. Pipes is a manager at KPMG US seconded to KPMG Belgium. She is specialized in privacy in a US and European context, and has experience with Big Data, Surveillance, social media, Internet of Things, (late) Safe Harbor, HIPAA/HITECH, COPPA, GLB and GAPP (Generally Accepted Privacy Principles).

D. Rotman is a managing director at KPMG US with over 25 years of experience. He is focused on providing Privacy, Security and Information Governance Services. He is the national Privacy service leader in the US.

C. Kypreos is a senior associate in the KPMG Cyber practice. He has more than six years of cyber experience and his primary focus is the delivery of Privacy and business resiliency engagements.

S. Cumming is an Advisory partner and head of Government Services for KPMG New Zealand. She has more than 30 years of experience, specializing in governance, risk management, project assurance, privacy and information management, health and safety, fraud and internal audit and assurance services.

A. van Kerckhoven is a manager at KPMG South Africa. She developed KPMG's Privacy By Design Methodology and is experienced in risk management, ISO 27001, PKI and security risk.

K. Hijikata is a director at KPMG Japan and a member of KPMG's global Data & Analytics team. He specializes in IT assurance, process mining, fraud detection, security and privacy. In the past, he has developed information systems and data warehouses.

T. Manchu is a senior manager in the Cyber Security practice at KPMG Japan, focusing on the security and privacy aspects at industrial and automotive companies. He has designed and implemented Information Security and Privacy Risk Assessment Frameworks in several companies.