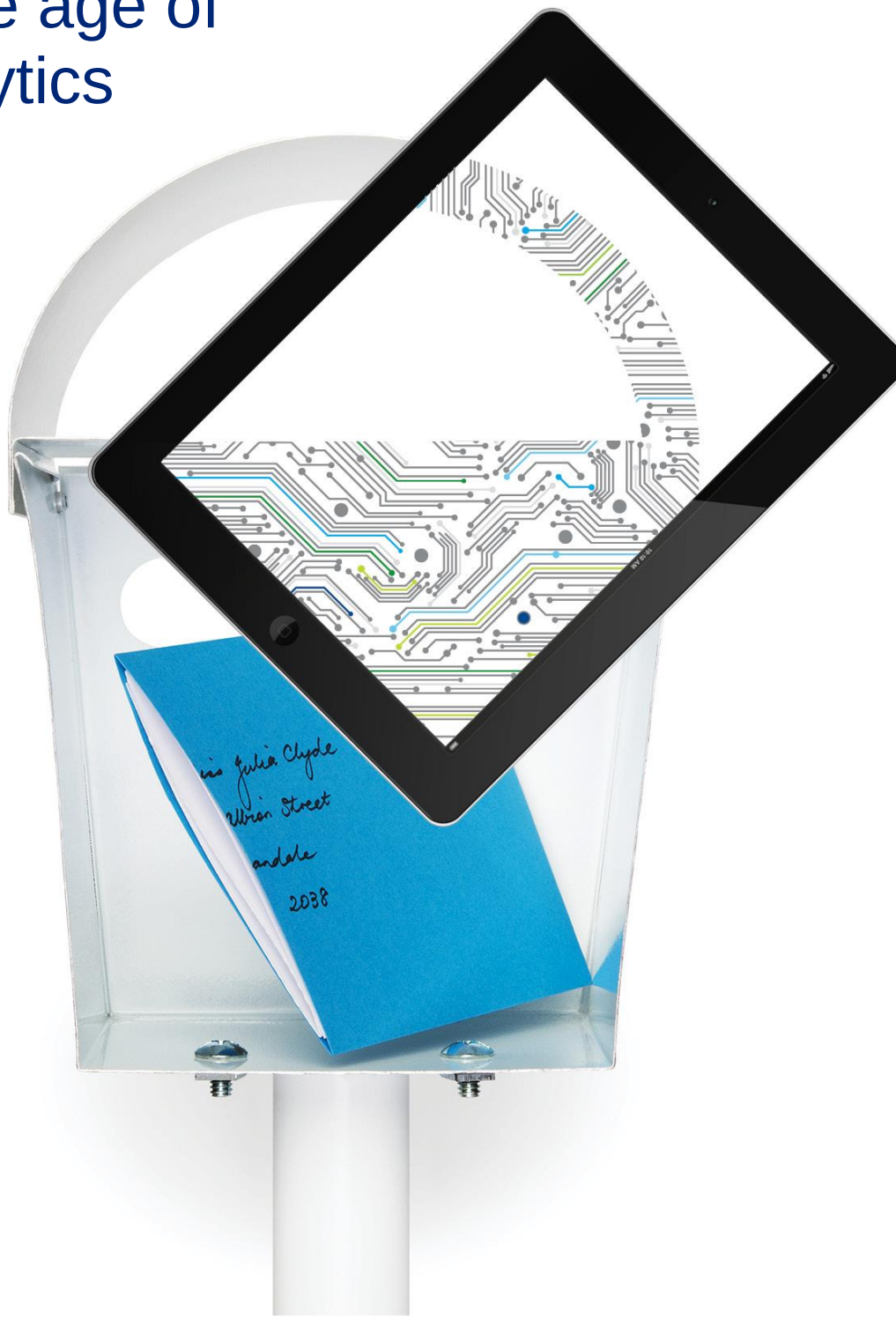
Deloitte.

Privacy by Design

Protecting privacy in the age of analytics



The era of Big Data is here, and it isn’t going away. The ability to use data to connect information, identify patterns and personalise interactions for maximum business result has reached an extraordinary level of sophistication.

Organisations will continue to use data analytics to advance their strategic goals, but the smart ones will embrace privacy as a driver of creativity and innovation and embed it into their systems to ensure quality results.

Through careful planning and application of privacy techniques and principles, organisations can use data to move business ahead and protect the personal information contained within them.

We can, in other words, **have it all**.

Executive summary

Big Data is about innovation

Data is undergoing a dramatic evolution. Businesses, governments and other organisations are unlocking value by turning everyday information into actionable insights. Two key trends are driving this shift:

- There's a lot more data. Every two days, the world creates as much new data as was created in all the years to 2003.
- With advances in data analytics, vast sets of both structured and unstructured data can be processed at very high speeds.

Our ability to connect information, identify patterns and personalise interactions for maximum result has reached a level of sophistication once only dreamed about.

Therein lies the problem.

Privacy is about personal information

Data analytics is so powerful it can combine data sets to infer someone's lifestyle, consumer habits, social networks and more than we can appreciate – even if no one data set actually reveals this personal information. We have technology that can take 1 000 words written by someone and provide a complete personal profile breakdown.

Not surprisingly, concerns are being raised over Big Data's impact on privacy. There are fears that fundamental protections, once taken for granted, are now challenged by the sheer velocity, veracity and volume of data and how it can be manipulated.

Some argue our very notion of privacy must change, that the imperative to innovate and unlock value from data must trump traditional concepts.

But this idea of trade-off between privacy and innovation is unhelpful and, frankly, outdated. We believe it is entirely possible to protect personal privacy while using data analytics to reveal new insights and innovation to advance progress. Indeed, just as technology gave rise to data analytics, it can also be used to solve the resultant privacy issues.

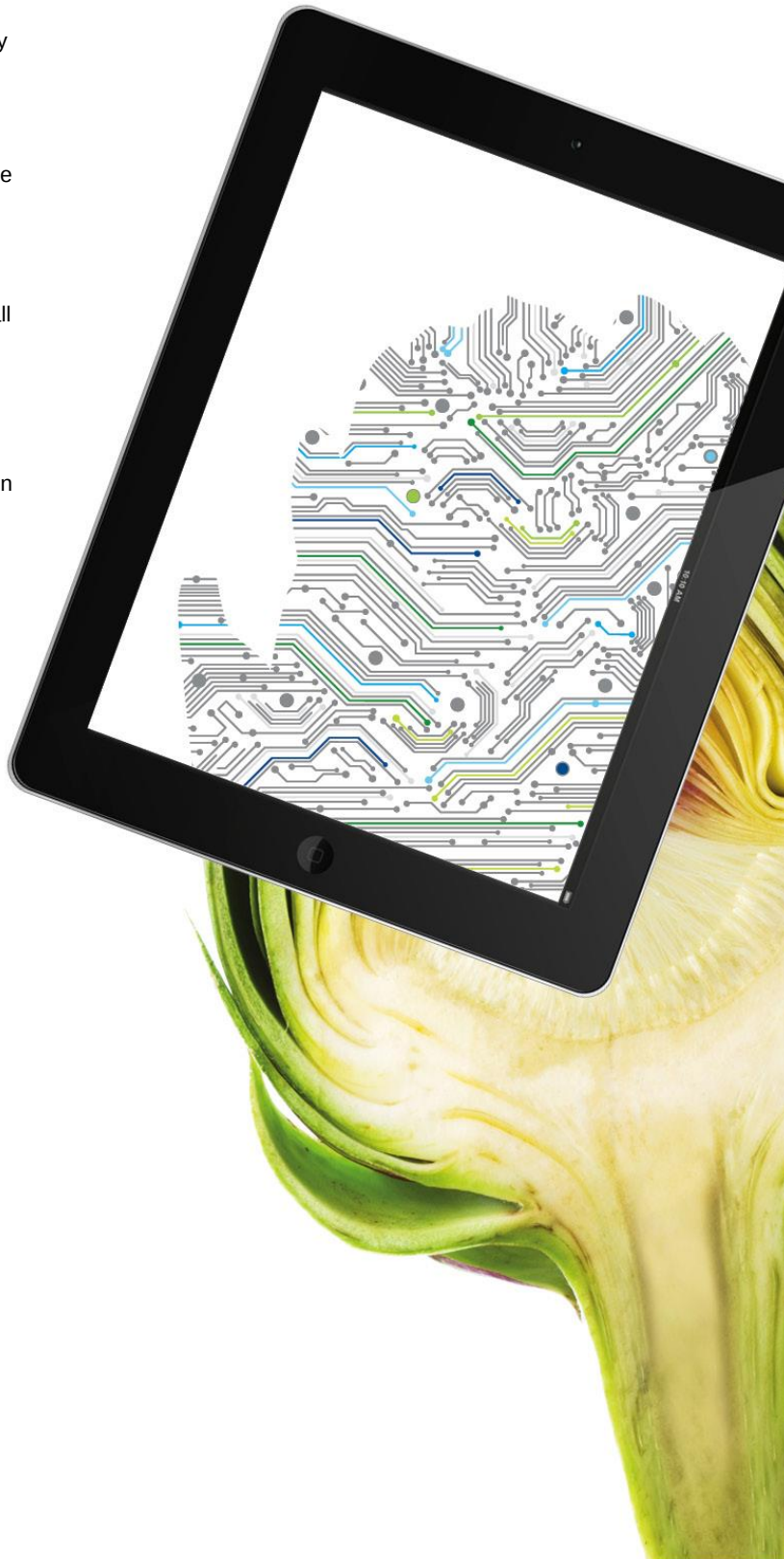
We can have it all

Privacy by Design (PbD) is designed to reconcile the need for robust data protection with the desire for data-driven innovation. Developed in the late 1990s by Dr Ann Cavoukian (Information and Privacy Commissioner of Ontario), PbD embeds privacy directly into the design specifications of technology, business practices and networked infrastructure.

Building on the PbD framework, several technology-based options for advancing privacy while pursuing data analytics are available to organisations, including:

- Data minimisation: No personally identifiable information is collected, unless a specific and compelling purpose is defined, all but eliminating privacy risk at the earliest stage
- De-identification: Datasets are stripped of all information that could identify an individual, either directly or through linkages to other datasets
- User access controls: A set of processes that grant or deny specific requests to obtain information; generally combined with other security policies

The Big Data era is here to stay. But that doesn't mean we must sacrifice privacy or shackle innovation. Through careful planning and application of privacy techniques and principles, such as those embodied in Privacy by Design, organisations can use data for its business needs while protecting the personal information contained in the data.



Big Data is big disruption

Big Data is about innovation

The amount of data (internet search queries, social media, GPS location information, stock transactions, etc.) generated by individuals, internet-connected devices and businesses is growing at an exponential rate. There are currently 9.6 billion internet-connected devices[1] and 1.3 billion mobile broadband connections[2] in the world. Every two days, these devices create roughly five exabytes (1018) of data. That's as much as all the data created by humans from the dawn of civilisation to 2003.[3]

The result is that, in this era of Big Data, data is among any organisation's most valuable assets. Analysing it properly can provide essential insights to help organisations develop strategy, manage risk and deliver growth and operational performance.

Organisations are understandably keen to unlock the business value of the data they have been collecting. They want to use this data to make smarter decisions that improve customer service, process efficiencies and outcomes.

And they can.

Recent and rapid advances make it possible to process large amounts of structured and unstructured data at very high speeds. Data analytics is accelerating the pace of innovation and disrupting traditional business models. For example:

- Retailers are tailoring their marketing to customers' preferences and purchasing behaviours.
- Financial services organisations are delivering advice and product recommendations before clients know they want them.
- Healthcare organisations are improving diagnoses, treatments and public health management.
- Government is making their data available to the public to increase transparency and encourage public engagement.
- In some industries, competitors are sharing data to address common concerns such as fraud, cyber security, and health and safety performance.

Put simply, today's data analytics enables organisations to make connections, identify patterns, predict behaviour and personalise interactions to an extent that could scarcely be imagined just a decade ago.

And therein lies the problem.

[1] IMS Research, "Internet connected devices approaching 10 billion, to exceed 28 billion by 2020," October 2012, http://imsresearch.com/ press-release/Internet_Connected_Devices_Approaching_10_Billion_to_exceed_28_Billion_by_2020&cat_id=113&type=LatestResearch.

[2] GSMA, http://www.gsma.com/newsroom/gsma-research-demonstrates-that-mobile-industry-is-creating-a-connected-economy.

[3] M.G. Siegler, "Eric Schmidt: Every 2 Days We Create As Much Information As We Did Up to 2003," TechCrunch, August 4, 2010, http:// techcrunch.com/2010/08/04/schmidt-data.

Big disruption equals big risks

In particular, organisations must be alert to threats of unauthorised access to data, especially personal data. More broadly, these risks can include reputational harm, legal action, regulatory sanctions, disruption of internal operation and weakened customer loyalty – all of which can result in revenue and profit losses.

The biggest risk that Big Data poses is the creation of automatic data linkages between seemingly non-identifiable data that can result in a broad portrait of an individual. A computer can achieve this by using 1 000 words that an individual has written, to create a full profile. It doesn't need to link datasets together or sneak or steal information.

Powerful analytics solutions can link data sets to reveal someone's lifestyle, consumer habits, social networks and more – even if no single data set reveals this personal information. A telephone number or postal code, for example, can be combined with other data to identify the location of a person's home and work. An IP or email address can be used to identify consumer habits and social networks.

Other examples of risk include:

- Authorised disclosure, loss or data theft is clearly a threat to privacy and is of course more serious when the Big Data set contains centralised, identifiable information. In extreme cases, unauthorised disclosure of personal information can put public safety at risk.
- Nudging is the use of identifiable data to profile individuals in order to analyse, predict and influence their behaviour. For example, someone with a bias against scarcity will automatically be served an ad that states "while supplies last", while a person with a bias for following others will get an ad labelled "best- selling". While it's gaining popularity, nudging may be perceived as invasive.
- Outsourcing data analytics can make managing accountability more difficult.
- Secondary use of data also raises concerns. In general, organisations can only use individuals' personal information for the purpose(s) identified at the time they collected the information with the person's consent. Using that information in analytics may be considered a secondary use, and unless the individual gave express consent, it can be seen as a breach.

The overriding issue here is privacy.

Privacy is personal

Information privacy refers to the right or ability of individuals to exercise control over the collection, use and disclosure by others of their personal information. Almost any information, if linked to an identifiable individual, can become personal.

And individuals are growing concerned:

- Ninety-three percent (93%) worry about their privacy online.
- Forty-five percent (45%) do not trust companies with their personal information.
- Eighty-nine percent (89%) avoid doing business with companies that they believe do not protect their privacy.

But not all data is personally identifiable, and not all non-personal data is the same.

- De-identified information refers to records that have had personal information removed or obscured, such that there is no reasonable basis to believe that the information can be used to identify an individual.
- Aggregated information is information whose values have been generated by performing a calculation across all individual units as a whole. For example, while uncovering new treatment strategies, medical researchers might look for patterns in aggregated patient data but have no way of identifying specific individuals.
- Non-personal, confidential information is information that often holds tremendous value and importance for organisations, such as business plans and proprietary research or other intellectual property. The disclosure or loss of such confidential information can be of grave concern for organisations. However, it does not constitute a privacy breach, because it does not involve the handling of personal information.

Some kinds of information are not so easily characterised. One example is metadata – information about other information, such as call lengths and other data about functional usage generated by mobile phones. As it happens, the detailed pattern of associations revealed through metadata can be far more invasive of privacy than merely accessing the actual content of one's communications.

Big data and privacy
are not mutually exclusive

Predictably, there are differing views. Some say big data analytics challenges fundamental privacy protections, while others argue that our privacy requirements are a barrier to the fruits of advanced analytics. But neither argument resolves anything. A new solution is needed – one in which the interests and objectives of both innovation and privacy can be met.

The view held in some quarters that privacy stifles innovation is as dated and flawed as the notion that privacy must be sacrificed for innovation. In fact, the opposite is true: privacy drives innovation because it forces innovators to think creatively to find solutions that can serve multiple functionalities.

We believe it is entirely possible to achieve privacy in the Big Data era while using data analytics to unlock new insights and innovations to move organisations forward. In our view, compliance-based approaches to privacy protection tend to focus on addressing privacy breaches after the fact. Instead, we recommend that organisations build privacy protections into their technology, business strategies and operational processes to prevent breaches before they happen.

Fortunately, there is already a framework for doing just that.

Privacy by Design

One of the most widely recognised approaches to proactive privacy is Privacy by Design (PbD). Dr Ann Cavoukian (Information and Privacy Commissioner of Ontario) developed this framework in the late 1990s, in response to the ever-growing impact of information and communications technologies and large-scale networked data systems. The PbD concept is to embed privacy measures directly into IT systems, business practices and networked infrastructure, providing a “middle way” by which organisations can balance the need to innovate and maintain competitive advantage with the need to preserve privacy. (See sidebar.)

Implementing this framework can result in changes to governance structures, operational and strategic objectives, roles and accountabilities, policies, information systems and data flows, decision-making processes, relationships with stakeholders and even an organisation’s culture. It’s no flash-in-the-pan theory, either. PbD has been endorsed by many public-sector and private-sector authorities in the United States and the European Union among other public bodies around the world who have passed new privacy laws. The Protection of Personal Information Act (POPI) was passed into South African legislation in November 2013. Subsequently, several organisations have taken steps to become POPI compliant. Additionally, international privacy and data protection authorities unanimously endorsed PbD as an international standard for privacy.

Adopting PbD is a powerful and effective way to embed privacy into the DNA of an organisation. It establishes a solid foundation for data analytics activities that support innovation without compromising personal information.

The seven principles of Privacy by Design

1. **Use proactive rather than reactive measures; anticipate and prevent privacy-invasive events before they happen.**
2. **Personal data must be automatically protected in any given IT system or business practice. If an individual does nothing, their privacy remains intact.**
3. **Privacy must be embedded into the design and architecture of IT systems and business practices. It is not bolted on, after the fact.**
4. **All legitimate interests and objectives are accommodated in a positive-sum manner.**
5. **Security is applied throughout the entire lifecycle of the data involved.**
6. **All stakeholders are assured that whatever the business practice or technology involved, it is operating according to the stated promises and is subject to independent verification.**
7. **Architects and operators must keep the interests of the individual uppermost by offering such measures as strong privacy defaults, appropriate notice and empowering user-friendly options.**

Strategies and tools to safeguard personal information

Building on the PbD framework, organisations have several technology-based options for advancing privacy while also using data analytics:

Data minimisation – reduce the data collected

According to this strategy, no personally identifiable information is collected unless a specific and compelling purpose is defined – effectively eliminating privacy risk at the earliest stage.

De-identification – make individuals less identifiable

With this strategy, datasets are stripped of all information that could identify an individual – either directly or through linkages to other datasets.

User access controls – restrict access

This is a set of processes that grant or deny specific requests to obtain information. It is generally combined with other security policies to safeguard personal information.

Several emerging technologies hold much promise for enabling privacy and utility to co-exist. These research-level examples include:

Differential privacy

Random “noise” is injected into the results of dataset queries to provide a mathematical guarantee that the presence of any one individual in the dataset will be masked. Software evaluates the privacy risks of a query and determines the level of noise to introduce into the result before releasing it.

Synthetic data

As long as the number of individuals in the dataset is large enough, it is possible to generate a dataset composed entirely of “fictional” individuals or altered identities that retain the statistical properties of the original dataset, while delivering differential privacy’s mathematical noise guarantee

For more information, please contact:

Dave Kennedy
Managing Director: Risk Advisory Africa
Tel: +27 11 806 5340
Email: dkennedy@deloitte.co.za

Werner Swanepoel
Africa Leader: Risk Advisory Data Analytics
Tel: +27 11 209 6664
Email: wswanepoel@deloitte.co.za

Cathy Gibson
Africa Leader: Risk Advisory Cyber Risk & Resilience
Tel: +27 11 806 5386
Email: cgibson@deloitte.co.za

Daniella Kafouris
Associate Director: Risk Advisory Legal
Tel: +27 11 209 8101
Email: dkafouris@deloitte.co.za

Derek Schraader
Director: Risk Advisory (Cape Town)
Tel: +27 11 806 5337
Email: dschraader@deloitte.co.za

Wesley Govender
Director: Risk Advisory (KwaZulu – Natal)
Tel: +27 11 806 5718
Email: wgovender@deloitte.co.za

Africa

Graham Dawes
Rest of Africa Leader: Risk Advisory
Tel: +254 20 4230000
Email: grdawes@deloitte.co.ke

Authors:

Dr Ann Cavoukian, Ph.D.
Information and Privacy Commissioner

David Stewart
National Advanced Analytics Leader Deloitte
Tel: +1 416-775-7484
Email: davstewart@deloitte.ca

Beth Dewitt
Manager and Privacy Specialist Deloitte
Tel: +1 416-643-8223
Email: bdewitt@deloitte.ca