BEST PRACTICE GUIDELINE: BIG DATA

A guide to maximising customer engagement opportunities through the development of responsible Big Data strategies.
# Table of Contents

**Preface**  
Jodie Sangster, Chief Executive Officer, ADMA  

**Foreword**  
Tony Davis, Partner, Quantium  
Jason Juma-Ross, Managing Director Australia & New Zealand, Accenture Interactive  

**Part One: Understanding Big Data**  
1. Defining Big Data  
2. Types of data and non-marketing applications  
3. Big Data in a marketing context  
4. Driving value from Big Data – marketing applications  
5. How to approach a Big Data project  

**Part Two: How to manage the governance issues of Big Data**  
1. The regulatory context  
2. How does privacy law apply to Big Data?  
3. The new definition of personal information  
4. Australian Privacy Principles (APP)  
5. Responsible Big Data business practices – check list  

**Summary of considerations for Big Data**  

**Acknowledgements**
Preface

Big Data. Some say it’s just a buzz word, others say it’s a game-changer. Whatever your perspective, one thing is for certain – data and the insights that it provides will dramatically alter the way that we do business now and into the future.

From a marketing perspective, Big Data provides marketers with new opportunities and efficiencies that have previously not been possible. It has the ability to significantly enhance relationships that companies have with their customers – moving to true personalisation, relevance and engagement. However, it is not easy! Many organisations in Australia and beyond are grappling with the complexities that Big Data brings – not only how to take the first step down the path to developing a Big Data strategy but also how to really gain meaningful insights from the infinite data available. It’s also essential that we get it right! With the new Australian Privacy Laws, companies that enter into the Big Data world need to ensure their strategies are the right strategies, keeping them within the boundaries of the law.

ADMA understands the value of Big Data and the crucial role it will play in the future of data-driven marketing and advertising. To assist members, we pulled together the leading minds in Australia to provide their viewpoints and expertise on Big Data, including where to start and how it can be used in ways to transform your business. Importantly, the Big Data Guideline also provides a viewpoint on how the new Australian privacy laws will apply and the compliance considerations that should be taken into account when collecting and using Big Data.

This has all been encapsulated in The ADMA Big Data Guideline and is intended to provide a starting point for companies and marketers that are ready to ride the Big Data wave… or at least understand the opportunities it presents for marketers and businesses alike.

I hope that you find the guide useful, informative and inspiring.

Jodie Sangster
Chief Executive Officer, ADMA
Foreword

Big Data is the new reality, it has been for years though the label is still fairly new. Data, mathematics and analytics has underpinned human endeavour in science, medicine and technology for centuries. Business has now grasped the opportunity data presents, the most recent catalyst being the ‘digital revolution’ which created a tidal wave of new, real-time data, revealing new dimensions of consumer preferences, shopping and social habits. So marketing is in the advanced guard in the data opportunity, championing consumer understanding and creating new propositions and business from these new data resources.

The opportunity is great, and our responsibilities as marketers are greater still. Ensuring that consumers think and feel positively about our brands and businesses is at the heart of engendering loyalty and underpins a mutually rewarding exchange of value. Whilst many consumers are completely comfortable with the data trail they create, and are increasingly recognising benefits from its use, many others are nervous about the implications of allowing organisations insights into their personal behaviours and preferences. Big Data practitioners are looking to be informed in relation to both the regulatory environment as well as ‘softer’ aspects of the data opportunity.

This guideline is one resource which provides some useful definitions, tips and thoughts on best practices to navigate this huge area of business innovation. The key is to understand and remain well within Government regulations and customer expectations… and then establish powerful processes to process raw data into new customer understanding and in turn amazing new commercial propositions.

Tony Davis
Partner, Quantium
Foreword

Talking to professionals about ‘Big Data’ today invariably invokes their buzzword detectors. The reaction is justifiable because many executives have lost patience with superficial polemics around petabytes and privacy. Accessing the potential of Big Data demands that we get past these memes and tackle the real complexity underlying the subject. The ADMA Big Data Guideline furthers the debate by first defining ‘Big Data’, then exploring how Australian marketers can derive value from it.

Executives I talk to would often rather remove ‘big’ from the term, returning to a discussion on just ‘data’. They recognise that size is not what matters. What matters is that the demand from businesses to engage with data and analytics has expanded rapidly, but organisations’ ability to access, interpret and govern that data has not. The first step is in recognising that data issues are as much a function of complexity, fluidity and skill shortage as they are of technology capacity. The Guideline hones this distinction and then explores real-world applications of data and analytics by marketers, both big and ‘small’.

Another theme I often encounter is value return from data. Value is a subjective term and to derive sustainable value from data and analytics, we need to look at it from multiple perspectives. ‘Big Data’ will have limited impact if it is focused purely in the domain of corporate value. Fortunately, many executives have now realised that the sort of value ‘Big Data’ can drive is more akin to what Umair Haque refers to as ‘thick value’. Creating thick value means simultaneous value for citizens, customers, organisations and stakeholders. Creating thick value requires a new look at traditional business and economic paradigms. For marketers, it means less about seeking out ‘consumers’ to ‘target’ and more about creating relevance. It means less about always being ‘on message’ and more about engaging in a range of authentic and meaningful conversations.

The Big Data Guideline takes a first step in navigating this path to value. After defining the field of play, it moves on to discuss how to approach a marketing data project by outlining some of the basics. A comprehensive review of data governance then provides a starting point for many organisations who are engaging with these complex issues. Although the guideline does not provide all of the answers, it does provide a starting point. Executives looking to leverage data and analytics to create long-term, meaningful value, should use it as a reference point from which to embark on their journey.

Jason Juma-Ross
Managing Director Australia & New Zealand, Accenture Interactive
PART ONE: UNDERSTANDING BIG DATA

1. Defining Big Data

In 2013 the volume of data being created by humans daily is 2.5 quintillion bytes, and 90% of the world’s data has been created in the past two years.¹ This phenomenon and the term used to describe it (‘Big Data’) has led to misconception of the key issues and opportunities. For example, a survey carried out by Harris Interactive on behalf of SAP in 2012 showed that only 25% of respondents agreed on a similar definition of Big Data (although nearly 76% viewed Big Data as an opportunity). Across the survey, definitions of Big Data varied widely:

- 28% defining it as the massive growth in transaction data
- 24% described it as new technologies that address the volume, variety and velocity challenges of Big Data
- 19% said Big Data refers to requirements to store and archive data for regulatory compliance
- 18% saw Big Data as the rise in new data sources, such as social media, mobile device and machine-generated devices.

So what does the term Big Data actually mean? As this paper will illustrate ‘Big Data’ is less about size and more about quality. Data volumes are often impressive by historical standards but technology advances are occurring so quickly that infrastructure is capable of accommodating the scale and speed of data processing required.

More challenging are the huge variety of data sources from mobile location data to call centre verbatims, and the fact that much of the critical data is streaming, operational information. Having the time to sit back and reflect on simple survey tables is often a thing of the past for today’s data scientists.

The defining dimensions therefore of what we know as Big Data are:

- Data from multiple sources
- Data in a wide variety of formats
- Data that is in constant motion
- Data volume can also be a feature.

The value derived from Big Data often lies at the intersection of two or more different data sets. Most of the ‘Big Data’ applications outlined in this document are examples of solutions to problems which require the distillation of selected elements from different data sources. These data sources may be unrelated, disconnected or un-matchable in their raw form – or a combination of all three. So complexity, driven by variety is a big problem.

To add to this complexity many Big Data applications involve analysing real-time data on the fly. For example, the monitoring and analysis of social media commentary around service outages helping to direct network remediation teams to the right location.

Increasingly, the focus in Big Data has shifted to commercial relevance and driving the value equation. Organisations that do this well have moved beyond the traditional ‘measure, analyse, report’ cycle to a much more integrated approach to Big Data. In today’s data powerhouses (such as Google, Amazon, Yahoo! and Netflix) ‘Big Data’ is in the blood and informs every decision and business process from the inside out.

In the graphic below, there is a global definition, and a subset of that – the commercial definition – beneath which is another subset – the marketing definition. The latter is more specific – elaborating on the applications in a general sense for Big Data. We’ll return to the marketing definition later. For now let’s concentrate on the broad definition.

### Definitions of Big Data

**ADMA Big Data Taskforce**

#### Global Definition

Big Data is the collection of large volumes of varied information, used to extend our understanding of the environment, medicine, science, business and human experience.

“A revolution that will change the way we live, work, think.”

#### Commercial Definition

Big Data is the current term given to the wide use of data collected from digital, technological, and analogue sources. Big Data is used to improve business understanding of markets, allowing improvements in customer experience and organisational performance.

#### Marketing Definition

Big Data is the current term given to collecting, analysing and generating insights from a wide variety of customer, commercial and environmental information.

It is used to develop better understanding of customer preferences, habits and considerations in making transactions with different categories, brands and channels.

The successful use of data in marketing leads to improved customer experience, a better exchange of value between customers and organisations, and improved business performance.

#### In short...

The collection, analysis and generation of insights from a wide variety of data sources in order to be able to improve business performance.
2. TYPES OF DATA AND NON-MARKETING APPLICATIONS

Types of Data

Below is a summary of some of the main categories of data typically regarded as Big Data:

1. Web behaviour and content:
The scope of web behaviour is huge. There are nearly five billion indexed web pages on the internet and for each page there are traffic statistics ranging from the number and duration of visits to far richer information on user behaviour on a large proportion of websites. Big Data also encompasses the content of those web pages and the changes that occur on them. Also included in this category is the vast amount of search engine data constantly being generated.

2. User-generated content:
This covers the content generated by the millions of social media users globally. The infographic below underlines the rate at which new content is being added – status updates, tweets, image uploads, video uploads, blog posts, forums, wikis etc. Email is also a massive content generator. Communications from the billions of mobile devices in use – both text, image and voice-based, also fall into this category.

Big Data – Velocity

Every 60 seconds

- 98,000+ tweets
- 695,000 status updates
- 11 million instant messages
- 698,445 Google searches
- 168 million+ emails sent
- 1,820TB of data created
- 217 new mobile web users

Challenges:
Handling real-time speed for updating, integrating, relevancy
3. RFID data:
RFID stands for Radio Frequency Identifier, which is a ‘tag’ commonly used for tracking purposes. There were an estimated 30 billion RFIDs in circulation by the end of 2011. RFIDs have a variety of uses from tracking the movement of products through a production process, to collection of tolls on a highway, to passports with embedded tags, to sporting performance to contactless payment. The type of data generated varies according to the application.

4. Location (or geo) data:
This can be generated in part by RFID tracking but a huge amount of location data is generated through the use of mobile devices. GPS now makes it possible to pinpoint a location to anything you do on your mobile, whether it’s making a call, taking a photo, or searching for something on the internet. Again, given the billions of GPS-enabled devices in circulation and the frequency with which they are used, this is constantly generating a huge amount of data.

5. Environmental data:
Well-established examples of this would be weather conditions, tidal movements, and seismic activity. It includes historical records as well as new data being recorded every minute. Just as mobile data has experienced massive growth in recent years, environmental data will be a major growth driver from wearables such as Apple bracelets, Google glasses, etc.

6. Organisation operational data:
Increasingly large amounts of data are continuously generated by organisations as a by-product of their normal activities, processes, transactions and interactions with the people who come into contact with them for whatever reason. This could be from public sector sources or commercial sources. It is generally structured and includes archived data as well as current data.

7. Research
This is another category with a very broad scope. Research includes government statistics on all manner of topics from population census data to research on health, education, law and order, economic activity, mining and agriculture. Research also covers the large volume and wide variety of statistics and information, both qualitative and quantitative, which is produced by research organisations and educational institutions on almost every topic of academic or commercial interest.

Big Data non-marketing applications
The first thing to point out is that Big Data is not Big Brother. The innovations required to sift-through, process, analyse and interpret this data represent a major technological milestone similar to the invention of the PC, the internet and the smart phone. It brings with it a new frontier of possibilities. Here are a few examples of how Big Data is enabling progress in a variety of positive ways outside of marketing.

1. Fighting crime with Big Data
In Los Angeles the police department have collaborated with the University of California to build a model from the millions of archived crime records, which helps them predict where and when crime will take place. This seemingly improbable mission uses a branch
of mathematical theory used in the prediction of earthquake after-shocks following a major earthquake. The model was tested in a suburb of LA, producing a daily hot-spot report for each police unit showing the areas where crime was most likely to occur. The police altered their routes to ensure they spent more time in those areas, deterring criminal activity by their presence. Over a period of time the test resulted in a decrease in property-related crime by 12% and a decrease in burglary by 26%. It is now being rolled out across LA and being tested in 150 cities across the USA.

2. Predicting outbreaks of disease
Data analysis can be beneficial in predicting the occurrence of disease. Researchers have been able to capture accurate data reflecting patterns of disease outbreak well before official government sources, by looking at online behaviour such as search queries and social media data such as tweets (delivered in high volume and at real-time velocity). In this category, a well-known example is the Google Dengue Trends research, which demonstrated how aggregate search query trends could provide an accurate, reliable model of real-world phenomena, such as outbreaks of Dengue fever (http://www.google.org/denguetrends/about/how.html).

3. Tackling emergencies
A non-profit tech company called Ushahidi used crowd-sourced data to enable more efficient emergency response policy development. Following the earthquake in Haiti, a centralised text-messaging system was set up to allow mobile phone users to report on people trapped under damaged buildings. Analysis of this data found a high correlation between the aggregated text messages and the concentration of damaged buildings, which demonstrated that the system could predict the extent and location of structural damage post-earthquake. This allowed more effective direction of resources, and maximised the benefits of those efforts.

4. Public sector economics:
A study carried out in 2011 by McKinsey & Co found that harnessing Big Data in the public sector has enormous potential. Their report concluded that if US healthcare practitioners were to use Big Data creatively and effectively to drive efficiency and quality, the sector could create more than $300 billion in value every year. Two-thirds of that would be in the form of reducing US healthcare expenditure by about 8 percent. Meanwhile, in the developed economies of Europe, government administrators could save more than €100 billion ($149 billion) in operational efficiency improvements alone by using Big Data, not including using Big Data to reduce fraud and errors and boost the collection of tax revenues.

2 Ibid, p23.
3. BIG DATA IN A MARKETING CONTEXT

In the previous section several non-marketing examples showed the kind of benefits which can be gained from harnessing Big Data. The potential to derive value from Big Data for marketers is huge, as underlined in a recent Gartner report which concluded that companies that are smart with data will have a 25-30% competitive advantage annually – those that don’t will die. However that potential to date is largely unfulfilled. For example, advertising efficiency is a major opportunity for improvement yet today, nearly 40% of advertising dollars are wasted, up from 37% five years ago. That makes for a $200B+ global wastage. Also 81% of CMOs list Big Data analytics as their number one priority yet 70% say they are unprepared for the Big Data explosion.

So how can Big Data improve marketing ROI exactly? To answer this question let us start by defining Big Data in a marketing context then (in the next section) we will examine a range of applications which demonstrate how organisations are using Big Data to improve both their marketing planning and the effectiveness of their marketing communications efforts.

We can view Big Data as harnessing data beyond an organisation’s ‘normal scope of operations,’ and extracting value from it. The following diagram helps to define ‘data beyond an organisation’s normal scope of operations’ by categorising all the data which could potentially be available for marketing purposes.

"a recent Gartner report concluded that companies that are smart with data will have a 25-30% competitive advantage annually"
Here it is represented as a series of spheres rather like a model of the universe with internal data – that which is within our normal scope of operations – sitting at the core, our online assets orbiting around that core with the greater universe of data beyond. Outside of this greater universe sit a number of organisations which specialise in making some sense of the data universe for the benefit of the marketing industry (as well as other functions).

**The core:**
This is an organisation’s internal data and the traditional focus of marketing activity for as long as data-driven marketing has been in existence. A typical scenario is a central CRM system which either provides all the IT solutions needed by the organisation, or which draws on a range of separate functional systems to create a single customer view. This system could contain data from financial systems, operational systems, service, sales and marketing. It’s fair to say that many, if not most organisations, are still battling with this particular challenge – how to re-organise themselves and their IT systems to achieve this single customer view, quite apart from the even greater challenge of harnessing data outside of the organisation.

**Owned online assets:**
This sphere contains the organisation’s websites and online communities linked to their social media sites. People who interact with an organisation’s websites or belong to their communities have some connection with the organisation even though in many cases it will be fairly tenuous. For example, usually only a small proportion of people in a community are actively engaged with a brand, and the majority of visitors to a website are totally anonymous. Unlike the core data, there is a limit to how much data can be harvested from the people who generate the data within this sphere above and beyond quantitative and qualitative analytics.

**The greater data universe:**
Theoretically, the greater universe of data could include any of the categories which were listed above, although in practice the categories with the greatest relevance to marketers will be web behaviour and content and user-generated behaviour and content (e.g. social media), followed by mobile, geo and research. Much of this data is constantly being generated and could potentially be used in real-time by monitoring and listening for the signals and sifting out the nuggets of potential value. Some of the data – research for example – could be static and archived.

**Data aggregators:**
The term ‘data aggregators’ encompasses a wide range of organisations with different services and missions. Trying to categorise them is difficult – some are essentially publishers, some provide marketing services or provide software and hardware platforms, and some hold large datasets and research which are available for marketing purposes. Some provide several of these functions. The characteristic they all share is their common aim of helping marketers (and others) to gather and selectively organise relevant data in the universe for the sole purpose of providing them with value in terms of insights, targeting tools and an enhanced customer experience. N.B. The organisations shown on the diagram are examples only and are not intended to be a complete list.
4. DRIVING VALUE FROM BIG DATA – MARKETING APPLICATIONS

Now that we have a definition of Big Data from a marketing perspective let us illustrate how it is being used or could be used to improve marketing effectiveness. For this it is useful to view examples in the context of the typical objectives facing marketers – planning, acquiring customers, converting sales and customer retention.

In the diagram below are eight examples of how organisations are using or could use Big Data. These are just a taste of the many ways in which an organisation could use Big Data to achieve their objectives. The columns show the business objective; the rows show the data categories being used. Beneath the diagram is an explanation of each example.

<table>
<thead>
<tr>
<th>Data aggregators</th>
<th>Online universe</th>
<th>Online assets</th>
<th>Internal data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Credit Card Branding</td>
<td>HP Data Management Platform</td>
<td>Behavioural targeting</td>
<td>Media attribution</td>
</tr>
<tr>
<td>Intent-based marketing</td>
<td>Website customisation</td>
<td>Loyalty programs</td>
<td>Onsite + CRM</td>
</tr>
</tbody>
</table>

1. Credit Card Branding

In 2011 a credit card launched two years previously by a prominent US bank was struggling to meet its acquisition targets. To get to the root of the problem they compared the acquisition performance of five brands – their own + four others, including the market leader, against a range of mostly publicly available data. This included credit card satisfaction studies, research on card holder acquisition and usage behaviour, Nielsen data on competitive media spend and a communications audit on each brand’s usage of channels and messaging content at each touchpoint.

The analysis showed that their weakness compared to other brands was their inability to deliver on certain promises around service and the customer experience. This prompted them to make positive changes to the service aspects of their business, alter their brand promise and reallocate media spend to strengthen their presence online. The combined effect of these measures enabled the bank to raise its acquisition rate by 13%.
2. HP Data Management Platform (DMP)
In 2012 Hewlett Packard set about improving its digital marketing, which it perceived was hampered by a lack of ‘actionable user insights.’ The data collection and processing of data from its digital marketing activities across 170 countries, involving thousands of HP marketers, its agencies and its partners, was disjointed and inconsistent. There were data silos, a lack of standardisation, and gaps in the data being collected throughout the company. This produced an incomplete picture both of online customer behaviour and the effectiveness of its marketing efforts, and it prevented shared learning.

Over a period of several months HP implemented a DMP from Blue Kai. The DMP solution standardises the collection of data from both media-side and site-side digital marketing activity, manages it, categorises it and analyses it. In the US the implementation of the DMP has resulted in big improvements in site performance; its operational efficiency is on the rise as smarter, more automated tools replace manual processes; its media campaigns are more sophisticated, and the insights it is gleaning from a 360° degree view of customer online behaviour is helping HP with strategic planning in all aspects of its marketing.

3. Behavioural targeting
This example of the use of Big Data is a practice that is now common in the targeting of online display advertising. Until relatively recently, the only way to target an online ad was to buy space on a publishing site, just as you would for a newspaper ad. Increasingly though, this blunt form of targeting is being replaced by behavioural targeting or OBA (online behavioural advertising). In essence, this is targeting to an individual user based on the data held by the publisher about that user.

Behavioural targeting is offered by the larger online publishing networks such as Mi9, Google, Yahoo, Fairfax and News Digital. The behavioural component of this data arises from cookies placed on a user's device, which then tracks and records their internet activity – where they are located, what they are searching for, which sites they visit, what they purchase online, etc. Cross-referencing this data with other data held by these organisations and their partners allows them to build a profile which can be used to target relevant ads when a user visits websites within a network.

The process of identifying the visitor, retrieving the profile, finding a suitable ad and serving it to the user takes less than half a second. The benefits to the advertiser include the improved targeting, the opportunity to serve a more relevant ad and reduced wastage through better control of duplication.

4. Media attribution
The well-known marketers’ lament “Half my advertising doesn’t work, I just don’t know which half” is more pertinent than ever, now that most campaigns use multiple channels and the consumer path to purchase is anything but linear. Many organisations are now trying to unravel this puzzle to understand the relative contribution of each channel in their media mix to their overall success, and the process has become known as media attribution.

A good example of this is Aussie Home Loans, which has invested a considerable amount in media attribution. The project started with an analysis of Aussie’s entire archive of historical marketing data to build a media attribution model. This was followed by tagging all of Aussie’s digital assets – effectively attaching sensors – which enabled them to gather the
correct data for ongoing analysis. Then the process began of recording the path to purchase which consumers took across all participating paid and organic channels.

Over time they were able to quantify each channel in terms of its role as an ‘introducer, influencer or closer.’ As a result Aussie could identify channels which were not effective, leading to a reallocation of >$500,000 of media budget to more effective channels. Further efficiencies allowed a reduction of media spend overall by 18% and a reallocation of these funds to other activities.

5. Website customisation using third-party data
This example is related to the behavioural targeting described above inasmuch as it uses the same profiling data and logic. Behavioural targeting as it applies to online display advertising sits beyond the borders of an advertiser’s owned digital assets. In some cases, organisations are now working with publishers and other online data aggregators to profile visitors to their own sites.

To understand this, imagine a situation where someone who has just bought a car or is about to buy a car visits an insurance company’s website to obtain a quote. The insurance company is able to retrieve a profile of this visitor from search activity data and web behaviour data held by its partners. The website experience can then be customised to the visitor. So, for example, if the profile indicates that this is probably a young person buying a small hatchback, the visitor could be presented with relevant content and perhaps an offer which matches his or her preferred type of car. The use of data in this way leads to an improved customer experience for the website visitor and increases the chance of a sale.

6. Search intent and site relevance
One of the primary aims of SEO specialists is to create site content which matches users’ search intentions. Search intentions are revealed not just from specific keywords and phrases but also from the trail of search terms people use as they attempt to refine their search. These patterns change frequently, adding another level of complexity. A number of specialist solutions have emerged to tackle this challenge. One such solution is the Web Relevance Engine (WRE) – the platform on which the services of US-based SEO specialists Bloomreach are based. “Your content discovered” is their catchphrase.

The WRE acts as a kind of bridge between searchers who can’t find what they want and marketers who struggle to be found. The basic problem is that consumer intentions change so rapidly and arise from so many potential sources, that it’s impossible for an organisation’s IT and website management teams to keep pace with it through manual editing of sites, so that people can easily find what they’re looking for.

Part of the solution is the collection and analysis of data from over 1 billion pages per day to decipher the language of web-wide demand and adapt to changes in consumer intent. The other side of the coin is the connection of an organisation’s pages together in a kind of mesh which is not constrained by the set navigation paths within a site, and which more closely matches what people are looking for. They can even create a new page dynamically based on the organisation’s site template and the products they have in stock to match what the customer is looking for. In effect, the organisation’s pages change automatically to match the newest search terms.
7. Customer loyalty programs

Loyalty programs are a well-established feature of marketing within many sectors. Many of these programs have moved up a gear from simply accumulating points which are eventually redeemed for discounts, to a range of much smarter initiatives: segmentation based on shopping basket analysis; cross-promotion of products to individual customers based on their individual purchase behaviour; alerts for customers’ frequently purchased items when they’re on special offer delivered to an app on their smart phone, and so on.

The scale of the larger of these programs puts them squarely in the Big Data category. And the data drives more than just customer engagement. In the retail space for example, the analysis of consumer behaviour is a valuable ingredient in a retailer’s business planning, and sharing this analysis with its suppliers helps those brands understand how shoppers buy in certain categories.

Take Woolworths Everyday Rewards for example which now has over six million members. Woolworths serves around 28 million customers a week, which generates a huge amount of data, processed, recorded, analysed and acted upon – and the need for speed is paramount. Although the program is based largely on structured transaction data, the scope is broadening. Data generated by its customer mobile app for example is more varied and includes location data. Bringing in data from the brand’s own social networking sites and website analysis or linking customer behaviour to media spend to conduct media attribution analysis all add new levels of complexity to the overall picture.

8. Linking customer data to web activity

One of the most commonly perceived uses of Big Data is the potential to somehow turn largely anonymous activity on the internet into marketing opportunities. If it were possible to isolate those behaviours which were relevant to a brand, engage with those people and find a way to serve them, marketers could unearth a gold mine. Well, now it is possible to a degree to do this, and the closer the online behaviour is to our own online assets the easier it is to do. A case study produced by Digital Alchemy for one of their clients demonstrates how it can be done.

The company in question sells a range of office products online. When a customer makes a purchase they first log-in, then place the order, but most of the time customers browse products without logging in so those interactions are anonymous. However, by tracking cookies on customers’ computers they can link the user to their transaction history even if they don’t log-in when browsing. If they view a product category but don’t purchase at the time, the company can send out a triggered follow-up email with a customised offer for those products. Companies using this technique to link web activity to identifiable customers have found this to be more successful than their normal targeting methods. In the DA case study, for example, the conversion rates were over 40% better.
5. HOW TO APPROACH A BIG DATA PROJECT

The previous section outlines several examples of how Big Data can be harnessed to improve marketing effectiveness. These examples are the exception rather than the rule. There is a gulf between those organisations on the one hand for whom data is the lifeblood – companies like Google and Amazon – and those organisations who are still battling with the basics of data management. In between are a large range of organisations at various stages of experimentation with new data sources, analytics, tools and skillsets.

What are Big Data ROI leaders doing differently? Recent research published in the Harvard Business Review, in which over 1200 executives were interviewed, reveals factors that differentiate leaders from laggards:

- leaders, on the average, outspent laggards on Big Data initiatives by more than a factor of three.
- leaders were more internet-centric. For example, leaders received 42% of their sales via the internet whereas laggards received just 29%.
- leaders generally see more potential in Big Data. They see it as affording them a better opportunity to improve services, marketing, sales, R&D, etc.
- leaders make more use of unstructured data, or digitised text, video, machine and other data that doesn’t easily fit into traditional databases, and traditionally has been hard for computers to analyse. They also make use of external data (data they don’t own) more aggressively.
- leaders place data analysts into a centralised team, removing them from business functions. This has two key benefits – freeing them from concentrating on a narrow silo focus and enabling them to share ideas, skills and techniques which generate creativity and synergy.

For businesses in the early stages of adoption, many of the biggest challenges faced are found in the ‘disconnects of marketing:’

- online disconnected from offline
- advertisers’ data disconnected from publishers
- awareness campaigns disconnected from performance marketing
- marketing disconnected from customer experience.

More specifically there are several obstacles preventing marketers from maximising the value which can be derived from Big Data:

- **disconnected data:** Not having the right data – most marketers have an incomplete view of customers. Their traditional source of marketing data – the CRM – usually holds a relatively thin slice of potential customer data.
- **disconnected insights:** Even if marketers had the right data, most are limited in their ability to transform the data into insights. Many organisations have access to a lot of data but don’t know what to do with it.
- **disconnected actions:** Even if marketers have the right insights, they are not necessarily actionable. Data and insights may sit in silos, which restricts potential synergies, shared learning and outcomes which benefit the wider organisation and its partners.
From a quick review of the different examples above, it’s clear that there are many ways to go about harnessing Big Data for marketing. It all depends on the nature of the marketing problem:

- the credit card launch is essentially a standalone piece of research involving the collation of data to understand patterns and produce insights to assist in strategic planning decisions.
- the Demand-Side Platform (DMP) is effectively a one-stop shop solution for collecting, standardising, managing and using digital data to drive better media-side targeting and site-side optimisation.
- a search intent-based marketing approach leverages digital intelligence to optimise marketing activities and solutions to match emerging customer wants in real-time.
- behavioural targeting involves a combination of technologies such as demand-side platforms, ad exchanges, ad servers and exchange servers.
- the Woolworths example is an exercise in managing and extracting value from the huge volumes and complexities of data generated mostly by its own internal operations.
- the linking of web activity to customer history is an exercise in customer-tagging limited to the organisation’s website and CRM.

It is therefore hard to pinpoint exactly the process for harnessing Big Data because there are such a variety of solutions. Also the extent to which each solution should be regarded as a project, approached in the same way that you might approach an IT project, varies according to the application. A comparison of some of the examples outlined demonstrates this point:

- behavioural targeting is definitely harnessing Big Data, but it’s just part of a solution offered by an online advertising network. There is no project management as such.
- search intent-based marketing requires a high degree of customisation for which project management is essential.
- the Everyday Rewards program is at the opposite end of the spectrum, as it is a core part of Woolworths business strategy, which occupies an entire division of a large corporate.

There are common themes running through them all however, and to better understand them it is useful to consider a more familiar challenge – the creation of a Single Customer View (SCV).

Until fairly recently data-driven marketing was largely confined to database marketing, whereby data about customers is collected and managed on a system and used to drive intelligent targeting and relevant messaging. To be as effective as possible it is desirable to create a SCV – a complete picture of a customer’s dealings with an organisation. For this to happen the organisation has to collate relevant data from several disparate systems into a single ‘source of truth’ from which marketing activity can be driven.
To plan for the creation of a SCV, you need to answer five basic questions:

- **objectives** – Why do you want a SCV? How is it going to be used and what exactly are you trying to achieve in terms of acquisition, conversion and retention?

- **data** – What data is required for effective marketing, where can it be gathered from the many sources and systems available and in what format?

- **processes** – how can this data be obtained, formatted correctly, cleaned, standardised and fed to the SCV? How is its accuracy maintained?

- **infrastructure** – what software and hardware is needed to obtain, process, manage and use the data and who will supply it?

- **analytics** – what types of analysis are required to drive effective one-to-one marketing? What type of reporting do we need? How do we ensure the feedback loops are in place?

These same five questions apply just as readily to a Big Data challenge:

1. **What do you want to achieve? For example:**
   - create customer and market insights to help you make a range of strategic decisions
   - streamline and standardise the collection of data from digital marketing to drive better targeting and messaging through digital campaigns and onsite activity
   - customise the experience for users visiting your website to drive more conversions
   - establish the usefulness of each media channel to help you to allocate media spend more efficiently
   - deliver more relevant content to more users through search engine optimisation
   - retain more customers, drive customer value and improve the customer experience.

2. **What data do you need?**
   - from your internal systems
   - from your online assets
   - from the wider internet and other sources
   - can it be obtained from a third party?
   - is this data qualitative or quantitative or both?

3. **What processes do you need to put in place?**
   - how will the data be captured?
   - what tagging is required of sites, pages, ads, users, etc.?
   - how will the data get to where it’s needed?
   - how do we maintain consistency, quality and accuracy?
   - how do we ensure we can obtain and process the data at the optimum speed?

4. **What infrastructure do we need?**
   - what type of hardware solution is required?
   - can it be managed internally or should it be hosted elsewhere?
   - what type of software solution is required?
   - who will supply it?
5. What analytics do we need?

- how do we turn the raw data into useful, actionable information?
- do we need to combine data from different sources?
- what types of analysis will drive the marketing decisions and actions we need?
- do we need special analysis tools to do this?
- what skills do we need to achieve this?
- where can we find those skills? Who do we need to work with?

The approach outlined above can be applied at a project level. At a strategic level other considerations take priority including the structure of the organisation, the recruitment of people with the right skills, the investment in software and hardware, the choice of partners to work with. These considerations are outside the scope of these guidelines.
PART TWO: HOW TO MANAGE THE GOVERNANCE ISSUES OF BIG DATA

Introduction

Big Data provides organisations with the means to better understand customer preferences, behaviours and expectations. Such tools can be used to maximise business efficiency, reduce costs for both business and consumers, improve products and services, and create a personalised customer experience.

However, most consumers don’t have a broad understanding of how Big Data is used, and fewer still understand its benefits. Fuelled by media hype, fears about loss of privacy may lead consumers to suspect that Big Data is synonymous with Big Brother.

Proponents of Big Data need to address legitimate concerns about its uses to build consumer trust. Lack of trust could lead to increased government intervention, which invariably means regulatory restrictions that stifle innovation.

So what can marketers do? Marketers can play a key role in ensuring that implementation of Big Data initiatives address privacy and security concerns; help consumers to gain a better understanding of how they benefit from Big Data, and enable consumers to control and make choices about the uses of their personal information.

The answer is three-fold:

1. **transparency** – being transparent about data practices is essential. It is not just telling the customer that data is being collected, but explaining how and why their interests are being protected, and giving them choice.

2. **customer-first** – when considering how you will use Big Data, put the customer first. Ask yourself: are you really providing a better experience, or will your activities be perceived as crossing the boundary between personalisation and improved service as the customer sees it, and ‘creepy,’ ‘spooky’ or just plain unexpected.

3. **consider your brand, not just the law** – just because the law allows something to be done, that doesn’t mean it should be done if it is not going to reflect well on your industry, company or brand. Prior to using data, consider whether your activity will have a positive or negative impact on your brand and act accordingly.

In simple terms, put the customer first, or they will put you last!
1. THE REGULATORY CONTEXT

For marketers, Big Data involves collecting volumes of data (which may or may not include the collection of ‘personal information’) from a variety of online and offline sources to gain a deeper understanding of consumers. Non-personal data may be used for analytical purposes at an aggregate level to determine trends, patterns or behaviours of broad customer segments. Such uses have minimal privacy implications. However, when:

- data is appended to an identified individual’s profile to infer and attribute additional characteristics about that individual, or
- data about an individual is brought together with other information about that individual, or
- de-identified information about individuals is made available to anyone who might be able to use other information to re-identify the individual, we have entered the realm of personal information protected by privacy law.

Marketers must have a clear understanding of the regulatory environment for data-driven marketing. The core legislation that applies to all organisations is Australia’s Privacy Act 1988. However, it is important to note that there are a range of other statutes and codes of practice regulating particular industries and particular types of data. For example, there are separate rules around using telecommunications data.

Australia’s Privacy Act 1988, establishes rules about how personal Information can be collected, used and disclosed which include:

- collecting personal information with consent or as reasonably anticipated by a customer following notice about the proposed collection
- notifying the individual about the purpose of collection (a limited and specified purpose)
- providing individuals access to personal information about them
- providing the right to correct inaccurate personal information, and
- ensuring that safeguards are in place to protect personal information from unauthorised disclosure or misuse.

New Australian Privacy Principles (APPs) will become effective from March 2014. They are briefly summarised in Part 4 of this section.
2. HOW DOES PRIVACY LAW APPLY TO BIG DATA?

‘Big Data’ is no different from any other form of data. What distinguishes Big Data is its volume, variety and the speed at which it is collected, processed and transmitted: if Big Data includes any ‘personal information’ then the same rules under the Privacy Act 1988 apply to its collection, use and disclosure as any other ‘personal information.’ If it is also sensitive information (such as health-related information) or telecommunications data or otherwise subject to sector-specific rules, these rules will also apply. Therefore, it is critically important, when developing and implementing a Big Data strategy that you consider and understand what sort of data is being collected and whether it will be considered personal information.

The new definition of ‘personal information’ in the Privacy Act refers to an individual who is ‘identified’ or ‘reasonably identifiable.’

‘Personal information:’ This is information or an opinion about an identified individual, or an individual who is reasonably identifiable. This includes (a) whether the information or opinion is true or not; and (b) whether the information or opinion is recorded in a material form or not.

This new definition of ‘personal information’ potentially captures more data than under the previous law. This may include, for example, data that is collated around a unique identifier that relates to a single individual but does not reveal the individual’s name.

Note that information you collect which is anonymous or de-identified may also be considered personal information if:

• you combine this anonymous or de-identified information with other information that makes the individual ‘reasonably identifiable,’ or
• you enable others to do so (for example, by disclosing anonymous or de-identified information without effective safeguards to prevent any re-identification of individuals).

A person is ‘reasonably identifiable’ when:

• you collect or hold anonymous or de-identified information that may be combined with other personal information (whether or not there is an intention to do so). Here you must treat that information as personal information from the time it is captured; or
• you disclose anonymous or de-identified information that may be combined with other personal information (whether or not the person to whom you disclose this information states any intention to do so).

In each case you need to implement technical and operational safeguards and protections in relation to handling of this anonymous or de-identified information about individuals to reduce any risk of re-identification of individuals (such as through data matching) until that re-identification risk is ‘remote’ or ‘low.’

If re-identification risk is higher than ‘remote’ or ‘low,’ the fact that the data is anonymised or de-identified form is irrelevant: the information is still protected as ‘personal information’ and the privacy law still applies.

The reason for this is that the requirement when considering whether information is ‘personal information’ is to consider all relevant circumstances as to the possibility of re-identification.
In assessing whether a risk of re-identification is ‘remote’ or ‘low,’ the draft Australian De-identification Paper states that de-identification “administered to a high standard, together with appropriate risk management strategies,” may be taken into account in determining whether the risk of re-identification has been effectively minimised. To confirm whether de-identification has been successful, the organisation should:

- apply the ‘motivated intruder’ test – this test considers whether a reasonably competent motivated person with no specialist skills would be able to identify the data or information. The specific motivation of the intruder is not relevant. It assumes that the motivated intruder would have access to resources such as the internet and all public documents, and would make reasonable enquiries to gain more information.
- look at re-identification ‘in the round’ – that is, assess the risk of any agency, organisation or member of the public identifying any individual from the data or information being released, either by itself or in combination with other available information or data.

Depending on the outcome of the risk analysis and the de-identification process, organisations may need to engage an expert to undertake an assessment of the anonymous or de-identified information to ensure the risk of re-identification of individuals is low.

Re-identification risk can be reduced by disclosing anonymised or de-identified information about individuals only on a limited access basis with robust safeguards in place before the data can be made available to others. These robust safeguards may be a combination of technical, operational and contractual safeguards. The safeguards appropriate in a particular data analytics deployment require a privacy and security impact assessment for that implementation and therefore will vary, but may include:

- use of trusted third-party arrangements
- use of pseudonymisation keys and arrangements for separation and security of decryption keys
- contractual limitation of the use of the data to a particular project or projects
- contractual purpose limitation, i.e. the data can only be used by the recipient for an agreed purpose or set of purposes
- contractual restriction on the disclosure of the data
- limiting the copying of, or the number of copies of, the data
- physical access limitations, such as separation of staff working on particular projects
- requirement as to training of staff with access to data, especially on security and data minimisation principles
- personnel background checks for those granted access to data
- controls over the ability to bring other data into the environment (allowing the risk of re-identification by linkage or association to be managed)
- contractual prohibition on any attempt at re-identification and measures for the destruction of any accidentally re-identified personal data
- arrangements for technical and organisational security, e.g. staff confidentiality agreements
- arrangements for the destruction or return of the data on completion of the project, and
- penalties, such as contractual ones that can be imposed on the recipients if they breach the conditions placed on them.
3. THE NEW DEFINITION OF PERSONAL INFORMATION

When is an individual ‘identified’ or ‘reasonably identifiable’?

Australian privacy law doesn’t provide specific examples or guidance on when a person is considered to be ‘identified’ or ‘reasonably identifiable’ from information about that individual. However, looking at international and Australian precedents, it is likely a person will be considered:

- ‘identified’ if the information can be associated with a specific individual
- ‘reasonably identifiable’ if the information does not of itself enable an individual to be identified BUT there is a greater than a low or remote risk that the information could be combined or used together with other information to enable an individual to be identified.

Whether an individual can be ‘identified’ or is ‘reasonably identifiable’ depends on context and circumstances, and the risk of the individual being identified. This is explained in further detail below.

Information relating to individuals (that is, information that is not aggregated to the point where possibility of re-identification of individuals is remote or law) is likely to be considered ‘anonymous data’ or ‘de-identified information’ if there is only a low or remote risk that the information could be combined or used together with other information to enable an individual to be identified.
The marketer’s challenge – ‘reasonably identifiable’

Big Data and analytics capabilities are bringing us closer to a world where information that was once considered anonymous can be easily combined with additional data sources making an individual ‘reasonably identifiable.’

The challenge for marketers dealing with Big Data is the collection of anonymous information that can be combined with other information (which might be general knowledge or matched data sets) thereby identifying the individual and effectively transforming ‘anonymous data’ or ‘de-identified information’ into ‘personal information.’

If the possibility of this transformation is other than ‘remote’ or ‘low’, the organisation must treat the information as personal information.

For example: cookie data has been collected by an organisation on an anonymous basis, attributed only to a unique device identifier or IP address (identifying the device, not the person using that device). The person using that device then logs into an internet site, providing his or her personal information. As the organisation can easily connect the cookie data with the individual, the cookie data has been transformed from ‘anonymous data’ into ‘personal information.’ Therefore, before the organisation can use the cookie data for any purpose, the organisation must comply with the requirements of the Privacy Act – providing the individual with notification and choice.

A recent example of this phenomenon was when researchers took large volumes of anonymised location data from a phone service provider and were able to identify 95% of the individuals in a group of 1.5 million people by overlaying the mobile phone data with other publicly available information.

Whether an individual is ‘reasonably identifiable’ from Big Data collected requires a consideration of the cost, difficulty, practicality and possibility that the anonymous Big Data information will be linked with other available information in such a way as to identify the individual.

While it may be technically possible for an organisation to identify individuals from anonymous or de-identified information it holds (by linking it with other information held by it or another entity), it may not be practically realistic. For example, logistics or legislation may prevent such linkages of information or an organisation may have set up security or firewalls that are assessed to be effective to prevent such a linkage happening. In these circumstances, individuals are not ‘reasonably identifiable’ – therefore, the anonymous or de-identified information is not considered ‘personal information.’
What to consider when the Privacy Act applies to Big Data?

Where Big Data strategies include the collection and use of ‘personal information’ (as explained above) the Privacy Act applies – here marketers need to be conscious of their responsibilities under the Privacy Act.

An overview of the responsibilities are outlined below and in Section 4. However, this is only provided as a guide and it is strongly recommended that organisations conduct a privacy impact assessment and implement a compliance program prior to implementing a Big Data strategy.

De-identification

Effectively de-identified information (so that the possibility of re-identification of individuals is ‘remote’ or ‘low’) ceases to be personal information, and its use is not restricted by the Privacy Act. In de-identified form, personal information may be used for purposes beyond that for which it was collected. It can be a means of extracting further value from personal information. Therefore, de-identification is a useful tactic for businesses employing a Big Data strategy.

There are a number of methods for de-identification. Examples include:

- removing or modifying identifying features such as a person’s name, address and date of birth.
- removing or modifying quasi-identifiers (for example, gender, significant dates, profession, income). Carefully consider whether quasi-identifiers should be de-identified and whether some may need to be retained for the information or data to continue to be meaningful and usable.
- combining data categories to re-cast data categories that contain small values that could assist in identification of individuals — for example, combining values for 18–24 year olds with values for 24–30 year olds into a single category of 18–30 year olds.
- manufacturing ‘synthetic data’, which can be generated from original data and then substituted for it, while still preserving the value of the original data.

It is good practice to use re-identification testing – a type of ‘penetration’ testing – to detect whether there are re-identification vulnerabilities. This involves attempting to re-identify individuals from anonymised data. There can be advantages in using a third-party organisation to carry out the testing, as it may be aware of data resources, techniques or types of vulnerability that you have overlooked.

Beyond legalities

The scope of the new Privacy Act to restrict uses of Big Data means that marketers should adopt a risk management approach to their Big Data strategies.

Marketers should consider minimising the amount of data they collect. If possible, they should use aggregate or anonymised data to obtain insights through analytics. A data minimisation approach will help businesses avoid collecting data for which they cannot justify a use, and reduce the risk to everyone in the event of a data breach.
To place customer considerations at the centre of this process, consider conducting ‘customer expectation’ surveys on the issues of data collection and use. This will help clarify what customers find acceptable.

Businesses need to weigh up the risks to their reputation, the risk of civil fines or other sanctions under the *Privacy Act*, and their relationship with customers when acting on insights gained through analytics. These considerations stand above and beyond legal requirements.

A customer-centric approach is useful for guiding businesses when deciding to act on insights gained through analytics. It will mitigate the risk of the customer objecting to the data use.

It is a business imperative to ensure data is accurate, and that insights derived from the data are not used in ways that could be harmful to the individual consumer.
## 4. AUSTRALIAN PRIVACY PRINCIPLES (APP)

### APP 1: Transparency

Open and transparent management of personal information.

Transparency is about ensuring consumers are aware of the data being collected and the purpose for which the data will be used, thereby allowing consumers to make informed choices about sharing their personal information. To satisfy the requirement for transparency, the information made available must be simple and easily accessible. Transparency is therefore not only about whether you say it – it is also about whether it is said in a way, and with sufficient prominence, that individuals are likely to see and understand it.

The more unusual or exceptional the data use, the greater the transparency required.

**Privacy policy:** Organisations must have a clearly expressed and up-to-date privacy policy that is reviewed regularly. This should be in clear, understandable English. Staff should understand what the privacy policy means in practice, and be trained to implement it. A privacy policy is a legal document – it is required in order to address dealings with personal information and an organisation is obliged to follow it.

APPs 1.4 and 5.2 set out what must be in a privacy statement. This includes:

- the kinds of personal information collected and held
- how such information is collected and held
- the purposes for which the entity collects, holds, uses and discloses personal information
- access and correction procedures
- complaint-handling procedures
- information about any cross-border disclosure of personal information that might occur, and
- any significant handling practice should be identified, such as if a company has specific information retention or destruction policies or obligations.

Privacy policies must be made available free of charge. This can be on your website, but you should also be prepared to provide it – upon request – in any other format.

**Documented compliance practices:** You should also have documented privacy compliance processes and practices that apply within your organisation. It is important to note that the Australian Privacy Commissioner has the power to ask you to present your compliance practices document. These should be developed in advance of 12 March 2014.

These requirements create particular challenges for data analytics. If personal information will remain associated with an identifiable individual through the course of analytics, or there is a real possibility that analytics results in respect of a particular but anonymised person may subsequently be matched back to information about that individual (re-identification) by you or someone else, then the analytics process and use of outcomes will need to be transparently explained. In addition, if sensitive information is being used in any way, consent of the individual will usually be required.
TRANSPARENCY AND BIG DATA

In the realm of Big Data, the business processes of data collection and use are themselves complex and unfamiliar to consumers. Big Data is often collected:

- as ‘anonymous data’ that later becomes ‘personal information’ as outlined in Section 3, or
- from third-party sources (rather than from the individuals themselves) therefore the consumer will often be unaware the information has been collected and will often be required to be informed as to that collection and any subsequent use.

If you collect or hold Big Data as ‘anonymous data’ but there is any real possibility that it will be linked with other data sources at a later date that transform it into ‘personal data,’ you need to:

- clearly outline your Big Data collection practices in your privacy policy
- ensure that all future uses of that data are clearly outlined in the privacy policy.

APP 2: Anonymity and pseudonymity

The law now requires that individuals have the option of dealing with a company anonymously or under a pseudonym in some circumstances. You don’t have to comply with this if it is not practical to do so, or if the transaction requires identification by law (or court or tribunal order).

When you are considering whether you require the name of the individual, consider the following:

- Can you provide the service using anonymised or de-identified information?
- What are the consequences of not capturing personal information? Will this affect your ability to engage with the individual?

It might not be practical to offer anonymity or pseudonymity if you need to:

- process a payment or require the name for a future service
- intend to send personalised marketing or communications to the individual
- de-duplicate records and maintain data accuracy.

However, if you can undertake the business process or provide the relevant service without identifying the individual, then you should endeavour to do so.
APP 3: Collecting personal information

You are permitted to collect personal information either directly from the individual or from a third-party (e.g. online data gathering) provided you have complied with the APPs and in particular rules relating to collection (APP 3), notification (APP 5) and direct marketing (APP 7).

The Australian Privacy Principles allow for collection of ‘personal Information’ (that is not ‘sensitive information’ such as health information or information about ethnicity, religion etc.) where the collection is ‘reasonably necessary’ for, or directly related to, one or more of an entity’s functions or activities, subject to notice.

Whether the collection is ‘reasonably necessary’ will be assessed from the perspective of a ‘reasonable person’, not merely from the perspective of the organisation collecting the information and proposing to undertake the activity.

In relation to collection, the following rules apply:

- collect only what is:
  - reasonably necessary for the purpose, or
  - related to one of your functions or activities
- only collect sensitive information with consent
- only use lawful and fair means for collection.
COLLECTION AND BIG DATA

It is important to note that:

- organisations may only collect personal information if it is necessary for a business function or purpose
- even if ‘anonymous data’ is being collected, it needs to be treated as ‘personal information’ if there is a real risk that it may be linked to other data at a later date such that it may identify an individual.

This is relevant in the world of Big Data where vast quantities of data can be gathered. In such instances, organisations should put in place practices that ensure they only collect personal information (or anonymous data that will become personal information) if they can show the information being collected is necessary for a current business function or activity.

Where possible, only collect personal information directly from the individual.

APP 4: Unsolicited personal information

This new rule says that if you receive personal information that you did not request, you should destroy or delete it unless you can show that you could have collected it under the rules for APP 3, Collecting Personal Information.

This would apply where a customer provides you with information that you have not requested. Such data should not be stored or used. Instead, it should be deleted.

APP 5: Notification

You must notify the individual before or at the time of, or as soon as practicable after, collection as to how and why you are collecting personal information about them.

APPs 1.4 and 5.2 should be read together to work out what should be in your notification statement. Among other things, it should include:

- contact details of the organisation
- whether information has been collected from a third party and the circumstances of that collection
- whether the collection of the personal information was required by law (and details about that collection)
- the purpose of the collection
- the main consequences to the individual of not collecting the personal information
- any other companies, or types of companies, that the personal information might be disclosed to
- complaint-handling and access/correction information in the organisation privacy policy
- that the privacy policy contains details of how to complain about a breach of the APP’s and how the company will deal with such a complaint
- information about disclosure to overseas recipients, and
- the countries to which such information may be disclosed.
When collecting personal information directly from the individual, the notification should be given at or before the time of data collection. If the personal information is being collected from a third party, the notification should be given as soon as practicable and at least at the first time this personal information is used, for example, to contact the individual.

The requirement to notify applies when organisations:

- collect personal information from an individual
- collect personal information from a third party, or
- append opinions or insights gained from anonymous or de-identified data, to an existing record of personal information. (This is particularly relevant in the world of Big Data.)

Use the following chart as a guide to notification practices:

### NOTIFICATION AND BIG DATA

This requirement is complex when it comes to Big Data.

It is important to keep in mind that notifications do not necessarily need to be provided before or at the point of data collection. Instead, there is flexibility built into the regulations that allow a notification to be given „as soon as practicable“. That means you can provide notification at the time of use (such as when personal information is actually used for marketing to the individual through a particular channel).

Therefore:

- if you collect anonymous Big Data and then append this to a specific customer record (thereby transforming it into „personal information“) you will be required to „notify“ the individual at the first point that you use the additional Big Data (i.e. as part of the first communication that uses the data)
- if you collect Big Data that includes „personal information“ from a third-party source, you will again need to notify the individual at the first point that you use this data, as above.

Notification should be included in the first communication with the consumer. You may choose to:

- list all the information within the communication, or
- include essential information within the communication and provide a link to additional information. It is important to note that if you include a link it must be prominent and make clear that it provides access to the information required in the „notification requirement.“ A link entitled „privacy policy“ is unlikely to be sufficient.
APP 6: Use and disclosure
You can use the personal information for the purpose it was collected, but you must not use or disclose it for any other purpose, unless:

- you have the individual’s consent
- the individual would reasonably expect the information to be used for another purpose
- the purpose is a related purpose.

BIG DATA AND USE
To comply with this requirement you must ensure that the intended uses and disclosures have been thoroughly considered in advance by your organisation and then have been clearly articulated in both (i) privacy policy, and (ii) the ‘notification’ statement.

APP 7: Direct marketing
An entity must not use or disclose personal information for the purpose of direct marketing unless they follow these rules:

Data collected from an individual

- the individual would reasonably expect the organisation to use the information for direct marketing*
- the organisation has provided a simple means by which the individual can request not to receive direct marketing, and
- the individual has not availed themselves of this means.

*Note: If the individual would not reasonably expect it to be used for direct marketing, and it is impractical to obtain consent, you must follow the rules for data collected from a third party.

Data collected from a third party

- the individual has consented to the use of information for direct marketing or it is impractical to obtain consent
- the organisation has provided a simple means by which the individual can request not to receive direct marketing, and
- in each direct marketing communication the organisation includes a prominent opt-out statement or otherwise draws the individual’s attention to this option, and
- the individual has not availed him or herself of this means.

Note that direct marketing is not defined and therefore is likely to mean any marketing that is directed to an individual using personal information.
Data collected from an individual

- The individual would reasonably expect the organisation to use the information for direct marketing,
- the organisation has provided a simple means by which the individual can request not to receive direct marketing, and
- the individual has not availed him or herself of this means.

Data collected from a 3rd party

- the individual has consented to the use of information for direct marketing or it is impractical to obtain consent,
- the organisation has provided a simple means by which the individual can request not to receive direct marketing, and
- in each direct marketing communication the organisation includes a prominent opt-out statement or otherwise draws the individuals attention to this option, and
- the individual has not availed him or herself of this means.

BIG DATA AND DIRECT MARKETING

APP7 has been designed to capture the widest possible range of current and future practices of data-driven marketers. Where Big Data is collected, attributed to a specific customer and then used for marketing purposes, this principle will apply.

Big Data is commonly gathered online from third-party sources. As a result, the requirements listed under ‘Data collected from a third party’ above will apply. Therefore, each marketing message (by whatever channel) sent to consumers involving Big Data must include an opt-out opportunity. It is essential that when Big Data is attributed to a specific customer record, the record is flagged within your database to ensure each marketing communication includes an opt-out opportunity.

Organisations need to give careful consideration as to HOW the opt-out is provided. The recommended approach is for organisations to include a link within marketing communications to a ‘communications preference centre’ where the customer can choose the type of communications they would like to receive and by which channels. This ensures that organisations are responding to consumers’ preferences and avoiding a blanket opt-out from all future communications.

APP 8: Cross-border disclosure

A cross-border disclosure is when personal information collected in Australia can be viewed by someone in another country. Organisations are now accountable for what happens to personal information they have disclosed to an overseas recipient.

It is important to note that a cross-border disclosure can occur within a group of companies – therefore it is important to conduct an audit of where personal information is being viewed and by whom.
Personal information stored in another country is also considered to have been ‘disclosed’ if the storage provider can access the data, therefore, those using cloud-based services need to treat such arrangements as cross-border disclosure.

An individual must be provided the same level of protection for their personal data if it is disclosed in another country, as they would have in Australia.

Organisations may only disclose personal information to a foreign country:

• if it is subject to a similar law or legally binding scheme as the APPs and there are mechanisms available to the individual to enforce that protection,
• if the organisation contracts with the foreign entity to provide the same level of protection as available in Australia, and
• if the individual is informed and consents in advance.

APP 10 – 13: Integrity of personal information
APPs 10 -13 relate to the integrity of personal information. These rules are about quality, security, access and correction.

Quality
• Organisations must take reasonable steps in the circumstances to ensure information collected, used or disclosed is accurate.
• Accuracy includes ensuring information is up-to-date, complete and relevant.

Security
• Organisations must take all reasonable steps to protect personal information. This includes adopting both physical and technological security measures.
• Personal information must be protected from misuse, loss, disclosure or unauthorised access.

Access
• Organisations must give individuals access to personal information about them on request.

Correction
• Organisations must correct personal information on request.
• If personal information is not to be corrected, reasons must be given for refusal, and information provided on how to complain.
5. RESPONSIBLE BIG DATA BUSINESS PRACTICES – CHECK LIST

The following checklist is designed to assist with developing or building upon your own Responsible Big Data Business Practices.

1. INFRASTRUCTURE AND OPERATIONAL CONTROLS

a) Culture, policies and practices

• Our organisation looks at all ways in which Big Data policies and practices affect our customers, employees and other stakeholders.
• Senior management understand and are accountable for our Big Data policies and practices and our compliance with them.
• We evaluate privacy and security risks involved in Big Data projects at the outset and maintain a proportionate level of ongoing risk review e.g. weigh up risk, intent and potential consequences of re-identification of anonymised and de-identified data – assess level of detriment to data subject if re-identification occurs and assess whether any threat (i.e. damage, distress or financial loss) to data subject is likely – if likely, put measures in place to make such a threat to the data subject unlikely i.e. further anonymisation, narrower access.
• We respond promptly to any allegation of breach of our Big Data policies and practices, and maintain ongoing review and assessments.
• Our protocols include: access restrictions, how data can be introduced into the organisation, disclosure restrictions, purpose limitations, encryption, risk assessments, restrictions on unauthorised re-identification of anonymised data, appropriate responses to a privacy breach incident.

b) People, training and awareness

• We have an accountability framework to ensure compliance with the organisation’s Responsible Big Data Business Policies and Practices.
• We ensure appropriate confidentiality agreements are in place with staff and contractors and usual background checks made.
• We ensure staff and contractors are aware of their responsibilities and the policies and procedures in place relevant to their use of Big Data.

b) Auditing

• We periodically review the on-the-ground business activities against our Responsible Big Data Business Policies and Practices.

---

3 For example, does one role have ultimate accountability? Is accountability delegated down to specific departments? This is key – a criticism that has been levelled at organisations following serious data-breach incidents is that there were adequate policies and procedures in place but no one was following them.

4 i.e on data security, data minimisation – whether via organisational wide training, bespoke training for particular roles, online resources or via other training and awareness means.
2. TECHNICAL AND SECURITY

- We ensure early assessment and implementation of end-to-end security measures.\(^5\)
- We have implemented encryption and key management protocols and systems.
- We have considered the need for layered security for handling sensitive data or risk of damage to vulnerable groups of individuals in the event of inappropriate disclosure.
- We assess whether certain projects or data sets require periodic re-identification testing or penetration or ‘Pen’ testing to identify re-identification vulnerabilities or gaps that need to be addressed to maintain integrity of the de-identified data.

We have clearly communicated data security policies and consequences for non-compliance.

3. CONTRACTUAL

- Due Diligence – we know the service provider or business associate that our organisation is passing data to.
- We ensure there is a contract in place prior to any transfer of data which imposes restrictions on what the third party can do with the data, requirements re: de-identification, security obligations, prohibitions on re-identification, arrangements for data upon completion of project, restrictions on subsequent disclosure, clear consequences and penalties for breach of contractual obligations.
- We are aware of and are able to comply with our contractual obligations when handling data for a third party.
- We monitor the third party’s compliance of their contractual obligations – the level of monitoring is based on the nature of the Big Data project.\(^6\) We identify proportionate levels of monitoring required via our risk assessment process.

We promptly address any contract violation but assess proportionality of response required i.e. escalation, request undertakings re: change of practices, implement closer monitoring, seek damages, terminate.

4. TRANSPARENCY

- Our Privacy Policy\(^7\) explains what our organisation does with personal data, and on an analytical front. It is accurate, user-friendly and transparent. It helps to explain the technology used and highlights any activities that are not standard.

---

\(^5\) i.e. internal and external – rather than adding as a bolt on at different stages of a project.

\(^6\) i.e. not all third-party data processing relationships require the same degree of monitoring.

\(^7\) There is a clear trend toward an expectation for simpler, shorter and easier to read privacy policies, that focus ‘upfront’ on what is unusual in a particular organisation’s use of personal information and not what is common. This all comes under a new regulatory catchword, ‘transparency’. A transparent privacy policy is one of a number of tools that an organisation can use to clearly show customers that its Big Data practices are ethically based and responsible, as well as legal.
SUMMARY OF CONSIDERATIONS FOR BIG DATA

- Transparency is the key to trust. Use your privacy policy and collection notices to develop consumer trust in your data collection practices.
- Businesses should comply with legal requirements in the collection and use of data, as per the APPs.
- Businesses should assess, beyond the legal requirements, whether their use of data will be within customers’ expectations.
- Businesses should use data and analytics in a responsible manner, and review their practices so that they ensure they are delivering benefits to consumers, not just the business.
- What is a responsible use of data and analytics will be determined by the circumstances, and the specific risks that any particular data use creates.
- Data security should be assessed on the basis of the kinds of information collected and used, and the relative risks associated with that.
- Businesses should consider the vulnerabilities of particular market segments, such as children, in their use of data and analytics.
- Be aware that if you breach privacy principles, or act out of line with consumer expectations, you will invite further government regulatory intervention.
ACKNOWLEDGEMENTS

ADMA gratefully acknowledges the contribution of the ADMA Big Data Taskforce and extends our thanks for their participation. Their effort, valuable knowledge and experience shared greatly contributed to this Best Practice Guide.

Tony Davis  Quantum  Chair, Big Data Taskforce
Jason Juma-Ross  Accenture  Vice Chair, Big Data Taskforce
Warren Billington  Acxiom  Taskforce
Rodney Rosmarin  CBA  Taskforce
Christian Bartens  Datalicious  Taskforce
Louis Keating  Ngage  Taskforce
Yvette Deerness  Sainty Law  Taskforce
Des Virana  SAS  Taskforce
Nick Adams  Telstra  Taskforce

ADMA also acknowledges the assistance as to privacy regulation of Big Data with early drafts of this Guide.

Peter Leonard  TMT Partner of Gilbert + Tobin
and Director of iappANZ