Big Data & Privacy: does your organisation need an ethical based approach?

The theme for this year's Privacy Awareness Week is 'trust and transparency'. The focus is on consumer and community trust that flows to organisations who handle personal information in clear and transparent ways.

As the law lags behind in rapid technology innovations, particularly in big data, AI, machine-learning and the Internet of Things (IoT), there is a growing discussion about the merits of an ethical based approach to data analytics.

This article considers why an ethical based approach can build trust and transparency with consumers, and why it should be part of good Information Governance, as a means of maximising the value of information derived from data analytics while minimising risks.

Big data

Big data describes the large volumes of data held by corporations and governments. Using analytics technology tools, insights and knowledge can be derived from the data. These insights can then be used to make informed decisions, for example, in the development of new or improved products or services providing a competitive advantage and ultimately delivering results to the bottom line.

In other words, there is big value in the information that can be extracted from big data for both the corporate and government sectors. In December 2015, the Australian Government released its Australian Government Public Data Policy Statement as part of the National Innovation and Science Agenda and recognised data as 'a strategic national resource that holds considerable value for growing the economy, improving service delivery and transforming policy outcomes'.

IoT, AI and big data analytics

A large contributor to the exponential increase in data held by organisations is the growth of the Internet of Things (IoT). Cisco has estimated that by 2020, the total volume of data generated by the IoT will reach 600 ZB per year and global IP networks will support more than 26 billion devices connected to the internet (up from 16.3 billion in 2015).

The IoT is the interconnection of devices to the internet and involves the transmission of data between the internet and those devices - e.g smart devices in factories, healthcare and smart buildings. Increasingly, it involves the connection of devices, i.e. 'things' and people in various aspects of their lives – e.g through fitness trackers, medical devices and Google Maps.   The IoT enables the seamless integration of our devices with the internet, and with it our data.

The UK Information Commissioner's Officer **(ICO**) March 2017 report, entitled 'Big data, artificial intelligence, machine learning and data protection' [at para 11] describes the connection between big data, AI and machine learning as follows: 'AI can be seen as a key to unlocking the value of big data; and machine learning is one of the technical mechanisms that underpins and facilitates AI. The combination of all three concepts can be called 'big data analytics'.

Privacy challenges for big data

A key concern about the IoT and big data are the challenges it presents to maintaining privacy of personal information, particularly when analytics and profiling are involved. This is because the collection of personal data may involve not only data that has been consciously provided by individuals, but also personal data that is recorded automatically (e.g tracking of online cookies), or derived from other data or inferred through data analytics.

A comprehensive list of privacy challenges in big data were identified in an International Working Group on Data Protection in Telecommunications report, 'Big Data and Privacy: Privacy principles under pressure in the age of Big Data analytics' as follows:

- **Re-use of data for new purposes -** that organisations which use collected personal data as a basis for predictive analysis must ensure that the analysis is compatible with the original purpose for collecting the data.
- **Data maximisation where the value of data is linked to potential future uses -** this can challenge the privacy principle that requires the processing of data must be adequate, relevant and not excessive for the purposes that have been defined and stated at the time of collection.
- **Lack of transparency -** lack of openness on how data is compiled and used may lead to consumers to decisions they don't understand or have no control over – for example, in relation the use of data by data brokers and analysis companies.
- **Compilation of data may uncover sensitive information -** that a compilation of bits and pieces of information, which may not be sensitive themselves, may generate a sensitive result – such as being able to predict a person's health.
- **Risk of re-identification -** through compilation of data from several sources, there is a risk that individuals may become identifiable from data sets which appear to be anonymous at first sight.
- **Security implications** - the challenges of the additional infrastructure layers needed to process big data and encryption of large data sets; data breaches may have more severe consequences when large data sets are involved; and organisations that acquire and maintain large sets of personal data must be responsible stewards of that information.
- **Incorrect data -** the risk that decisions are based on inaccurate information, particularly when information is obtained from online sources and not official registries – for example, credit agencies.
- **Power imbalance -** between those that gather the data (large organisations and states) and individuals.
- **Determinism and discrimination -** because algorithms are not neutral, but reflect choices, among others, about data, connections, inferences, interpretations, and thresholds for inclusion that advances a specific purpose. Big Data may consolidate existing prejudices and stereotyping, as well as reinforce social exclusion and stratification.
- **Chilling effect** - this is the effect that people will restrict and limit their behaviour if they know or think that they might be surveilled.
- **Echo chambers** - which may result from personalised advertising, search results and news items so that people will only be exposed to content which confirms their own attitudes and values.

Big Data – Transparency & Trust

The lack of transparency, as identified in the above report, in addition to trust, are key issues for consumers when it comes to providing personal information to organisations and can provide an important competitive differentiator.

A Harvard Business Review **(HBR)** article in 2015, titled Customer Data; Designing for Transparency and Trust, referred to numerous studies having 'found that transparency about the use and protection of consumers' data reinforces trust'. To assess this affect, the authors carried out their own international survey to understand consumers' attitudes about data and what they expected in return for providing it. The authors conclude:

> 'A firm that is considered untrustworthy will find it difficult or impossible to collect certain types of data, regardless of the value offered in exchange. Highly trusted firms, on the other hand, may be able to collect it simply by asking, because customers are satisfied with past benefits received and confident the company will guard their data. In practical terms, this means that if two firms offer the same value in exchange for certain data, the firm with the higher trust will find customers more willing to share.'

The potential difficulty of collecting personal information is supported by the Office of the Australian Information Commission's longitudinal surveys into community attitudes to privacy, which reveal Australians are increasingly conscious of personal data issues. Privacy Commissioner, Timothy Pilgrim said in a speech to CeBit in May 2016:

> 'The majority of Australians – 60 percent – have decided not to deal with an organisation due to concerns about how their personal information will be used. And significantly, 97 percent of Australians don't like their personal information to be used for a secondary purpose. This is critical to big data. Because big data projects will often involve secondary use of data.
>
> If that data finds its source in personal information, then we have a clear dissonance between our known and understandable desire that our personal information works for us and for the purposes we explicitly provided it for versus the demonstrable innovative power of that data to improve our services and lives.'

The ICO 2017 report, titled 'Big data, artificial intelligence, machine learning and data protection', refers to ICO-commissioned research that shows 'the more people trust businesses with their personal data, the more appealing they find new product offers such as smart thermostats and telematics devices in cars'.

The HBR article and the ICO report support a business case for developing an approach that aims to build trust and is based on transparency and fairness.

How can big data be used in ways that respect the privacy of individuals?

An approach aiming to build trust and based on concepts such as fairness and respect, go beyond legal compliance and support an ethical based approach. Ethical considerations in relation to processing of personal information in big data are the subject of increasing global discussion and developments.

The Council of Europe's Consultative Committee Convention 108 in January 2017 issued 'Guidelines on the Protection of Individuals with regard to processing of personal data in a world of Big Data'. The first principle of the Guidelines is the ethical and socially aware use of data. It stipulates that in the processing of personal data, controllers should adequately take into account the likely impact and the broader ethical and social implications, and that it should not be in conflict with the ethical values commonly accepted in the relevant community or communities and should not prejudice societal interests, values and norms, including the protection of human rights.

In Australia, the NSW Data Analytics Centre was established in late 2015 to become a leader in whole-of-government data analytics to provide insights into complex policy problems and improve service delivery for the community. To help support this work and address the larger challenges of data sharing, a Data Sharing Taskforce was established in 2016 with participants from Australian peak bodies such as the Australian Computer Society, state and commonwealth governments, and private sector representatives to address the overarching challenge of developing privacy preserving frameworks which support automated data sharing to facilitate smart services creation and deployment. The Taskforce's report, expected to be delivered within the next 6 months, will identify best practice, models and frameworks for data sharing in the Australian privacy context.

In the United States, The Information Accountability Foundation **(IAF)** has been working on a Big Data Ethics Initiative since late 2014. The IAF's goal is 'to achieve effective information governance systems to facilitate information-driven innovation while protecting individuals' rights to privacy and autonomy.'

The IAF sets out the ethical data use principles in the paper**, '**Decisioning Process, Risk-Benefits Analysis Tool for Data Intensive Initiatives - Achieving Legal, Fair and Just Use of Data & Appropriate Individual Engagement'. The ethical principles articulated indicate that data use should be Beneficial, Fair, Respectful and Just, Transparent and Autonomy Protecting and performed with appropriate Accountability with a Redress Provision. The considerations for each of these ethical principles as proposed by the IAF are set out in the table below.

BENEFICIAL

- Uses of data should provide benefits and value to individual users of the product or service. While the focus should be on the individual, benefits may also be accrued at a higher level, such as groups of individuals and even society as a whole.
- Where a data use has a potential impact on individual(s), the benefit should be defined and assessed against potential risks this use might create.
- Where data use does not impact an individual, risks, such as adequately protecting the data, should be identified.
- Once all risks are identified, appropriate ways to mitigate these risks should be implemented.

FAIR, RESPECTFUL AND JUST

- The use of data should be viewed by the reasonable individual as consistent, fair and respectful.
- Data use should support the value of human dignity – that individuals have an innate right to be valued, respected and to receive ethical treatment. Human dignity goes beyond individual autonomy to interests such as better health and education.
- Entities should assess data use against inadvertent, inappropriate bias or labelling that may have an impact on reputation or the potential to be viewed as discriminatory by individual(s).
- Data should be used consistent with the ethical values of the entity.
- The least data intensive processing should be utilized to effectively meet the data processing objectives.

TRANSPARENT AND AUTONOMY PROTECTION (Engagement and participation)

- As part of the dignity value, entities should always take steps to be transparent about their use of data. Proprietary processes may be protected but not at the expense of transparency about substantive uses.
- Dignity also means providing individuals and users appropriate and meaningful engagement and control over uses of data that impact them.

ACCOUNTABILITY AND REDRESS PROVISION

- Entities are accountable for their use of data to meet legal requirements and should be accountable for using data consistent with the principles of Beneficial, Fair, Respectful & Just and Transparent & Autonomous Protection. They should stand ready to demonstrate the soundness of their accountability processes to those entities that oversee them.
- Individuals and users should always have the ability to question the use of data that impacts them and to challenge situations where use is not consistent with the core principles of the entity.

Ethical framework considerations

The need for an ethical based approach and the formality of the process will depend on the context, such as the size of the data set or sets and the amount of personal data to be processed within them. Whether a formal ethical framework or an ethics committee should be established will depend on the type and recurrence of data analytics being carried out and will vary across industries as well as government departments.

A comprehensive approach would include an internal ethics committee for the collection of data and data analytics. The ICO report into 'Big data, artificial intelligence, machine learning and data protection' points out [at para 176] that 'a large organisation may have its own board of ethics, which could ensure that its ethical principles are applied, and could make assessments of difficult issues such as the balance between legitimate interests and privacy rights'. Universities have a long history of research ethics committees and there are examples of medical and health organisations with an ethics based approach to the collection of data and data analytics. For example, the Centre for Epidemiology and Biostatistics at the Melbourne School of Population and Global Health developed the Guidelines for the Ethical Use of Digital Data in Human Research.

In the EU, the 'Guidelines on the protection of individuals with regard to processing of personal data in a world of Big Data' recommend the use of ethics committees where it is assessed there is likely to be a high impact of the use of big data on ethical values, as a means of identifying specific ethical values to be safeguarded in the use of data. The Guidelines at 1.3 provide that the 'ethics committee should be an independent body composed by members selected for their competence, experience and professional qualities and performing their duties impartially and objectively.'

As the ICO report points out [at para 176], an important issue 'is the organisational relationship between the ethics board and employees with responsibilities for data and analytics, such as the chief data officer and the data protection officer'. This highlights the importance of an overarching Information Governance approach between those managing the data, those responsible for its protection – i.e the chief privacy officer and legal teams, those performing the data analytics and those using the information obtained to ultimately derive value from it.

One of the key challenges facing organisations in the information age is the ability to take an organisational strategic approach to Information Governance to effectively maximise the value of information through data analytics while minimising its risks. These include the costs arising from:

- Data breach of personally identifiable information;
- Other breaches of privacy or other applicable legislation;
- Compliance with mandatory data breach notification requirements;
- Responding to regulators;
- Responding to and dealing with adverse publicity arising from perceived misuse of consumer information;
- Costs arising from an actual or perceived loss of trust by consumers and negative impacts, such as, ongoing damage to reputation and/or a negative impact to the bottom line.

Information Governance, privacy frameworks and the management of data analytics initiatives need to be aligned with overall strategic objectives led from the top down, comply with the law, and where technology is lagging behind the law, an ethical based approach should be used. An ethical based approach should be embedded into the organisation to enable employees working on data initiatives to have a clear mandate to apply guiding ethical principles or an ethical framework.

The way in which an ethical based approach may be embedded into an organisation will vary according to the types of data initiatives being undertaken, the volume of personal information involved or the potential for personally identifiable information.

Types of ethical based approaches include:

- Ethical value or policy statement for data initiatives, which can be used as a reference point for employees in any data initiatives to guide decision making and data impact assessments;

- Ethical frameworks and/or checklists for use in data initiatives – see for example: Data Impact Assessments including ethical aspects below; and Part B in the Guidelines for the Ethical Use of Digital Data in Human Research, which sets out five categories of ethical issues and guiding questions when conducting research involving digital data covering: consent, privacy, ownership, data governance, and data sharing; and

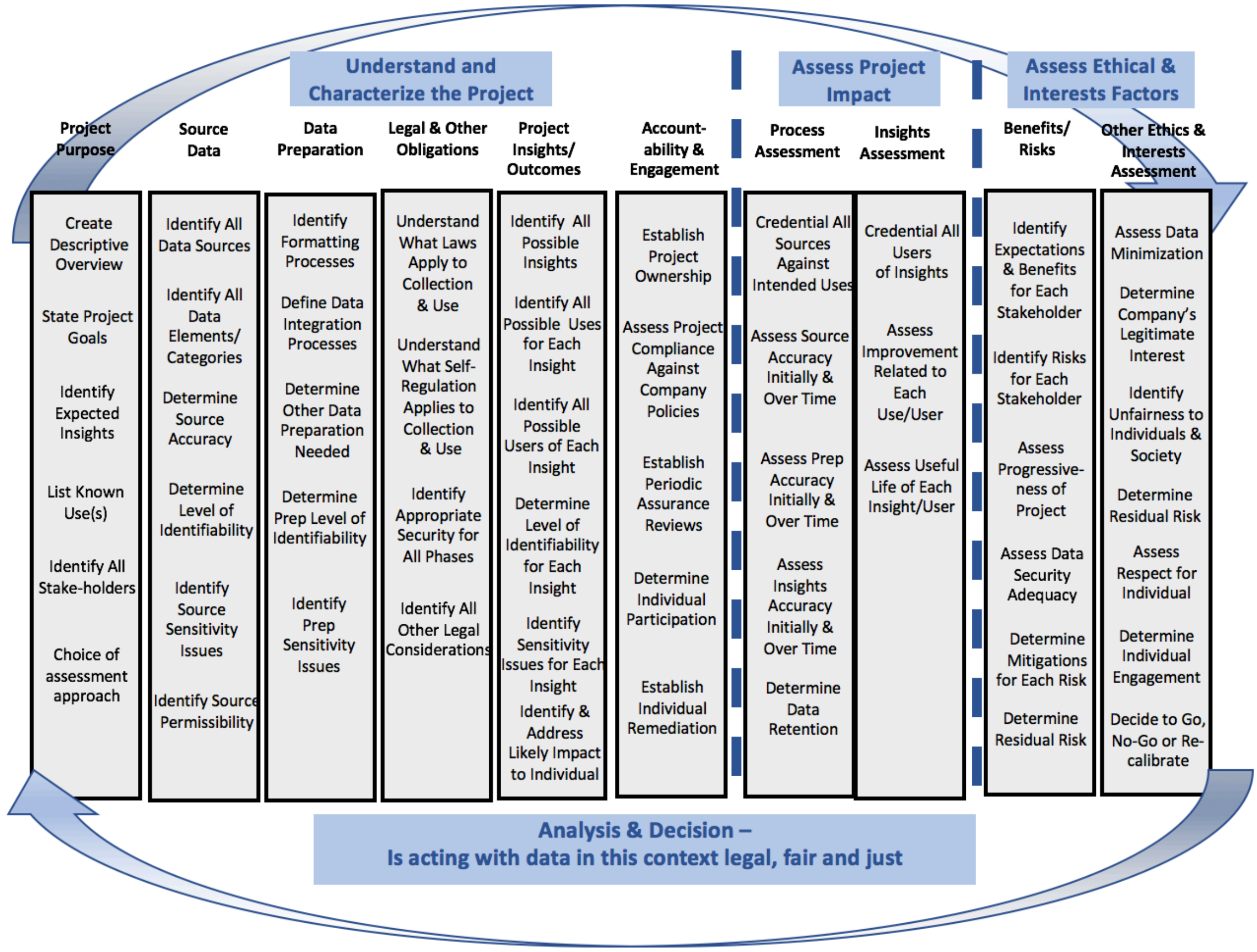
- Ethics committees or internal ethics boards.

Big Data Impact Assessment

As the IAF paper sets out whether the project is a core product review, the broader use of information, or a big data analytics project, an assessment process is required to address the legal, ethical, fair and other implications of information use.

The IAF paper proposes a Comprehensive Data Impact Assessment, as set out in the figure below, for data intensive initiatives to adequately and systematically (operationally) determine interests/impact to stakeholders, including specifically the individual.

CDIA for Data Intensive Initiatives



Understand and Characterize the Project

Assess Project Impact

Assess Ethical & Interests Factors

Project Purpose
- Create Descriptive Overview
- State Project Goals
- Identify Expected Insights
- List Known Use(s)
- Identify All Stake-holders
- Choice of assessment approach

Source Data
- Identify All Data Sources
- Identify All Data Elements/ Categories
- Determine Source Accuracy
- Determine Level of Identifiability
- Identify Source Sensitivity Issues
- Identify Source Permissibility

Data Preparation
- Identify Formatting Processes
- Define Data Integration Processes
- Determine Other Data Preparation Needed
- Determine Prep Level of Identifiability
- Identify Prep Sensitivity Issues

Legal & Other Obligations
- Understand What Laws Apply to Collection & Use
- Understand What Self-Regulation Applies to Collection & Use
- Identify Appropriate Security for All Phases
- Identify All Other Legal Considerations

Project Insights/ Outcomes
- Identify All Possible Insights
- Identify All Possible Uses for Each Insight
- Identify All Possible Users of Each Insight
- Determine Level of Identifiability for Each Insight
- Identify Sensitivity Issues for Each Insight
- Identify & Address Likely Impact to Individual

Account-ability & Engagement
- Establish Project Ownership
- Assess Project Compliance Against Company Policies
- Establish Periodic Assurance Reviews
- Determine Individual Participation
- Establish Individual Remediation

Process Assessment
- Credential All Sources Against Intended Uses
- Assess Source Accuracy Initially & Over Time
- Assess Prep Accuracy Initially & Over Time
- Assess Insights Accuracy Initially & Over Time
- Determine Data Retention

Insights Assessment
- Credential All Users of Insights
- Assess Improvement Related to Each Use/User
- Assess Useful Life of Each Insight/User

Benefits/ Risks
- Identify Expectations & Benefits for Each Stakeholder
- Identify Risks for Each Stakeholder
- Assess Progressive-ness of Project
- Assess Data Security Adequacy
- Determine Mitigations for Each Risk
- Determine Residual Risk

Other Ethics & Interests Assessment
- Assess Data Minimization
- Determine Company's Legitimate Interest
- Identify Unfairness to Individuals & Society
- Determine Residual Risk
- Assess Respect for Individual
- Determine Individual Engagement
- Decide to Go, No-Go or Re-calibrate

Analysis & Decision –
Is acting with data in this context legal, fair and just

Assessment of ethical aspects

The IAF propose that the following considerations should be considered when assessing the ethical aspects of a big data initiative:

- What aspect of collection/acquisition/processing/analysis or use of the insights could be considered unfair to the individual or to society?
- Is the collection/acquisition/processing/analysis or use of the insights done in a way that is respectful to the individual?
- Has the minimum possible amount of data been used?
- Does the company have a legitimate interest in the processing of the data and the use of the insights?

- After all mitigations have been applied, what is the residual risk to all stakeholders, particularly the individual – have the benefits and risks been effectively balanced?
- Have the interests, expectations and rights of individuals been effectively addressed?
- What additional, contextual based participation and choice (meaningful) with the individual should be considered?
- Is there an effective redress option for the individual impacted? Has this use of data been transparent?
- After considering all the above factors, is the project a 'go', 'no-go' or should some aspect be recalibrated to reduce the residual risks?

Conclusion

Organisations should implement strong Information Governance including privacy frameworks that include data impact assessments and an ethical based approach where data analytics of big data involving personal information is being undertaken.

The benefits of an Information Governance framework enable an organisation to take a strategic approach to both maximise the value of information derived from data analytics as well as minimise the risks arising from the costs of legal and regulatory privacy compliance, costs arising from data breaches and/or responding to regulators, as well as costs arising from loss of reputation, particularly where there is a breach of trust with consumers.

The type of ethical based approach, such as, whether it is an ethical value or policy statement, or an ethics committee, and formal data impact assessments including an ethics assessment will depend on the types of big data initiatives being undertaken, the volume of personal information involved or the potential for personally identifiable information.

An ethical based approach will ensure that the processing of personal information in big data analytics is carried out in a fair, transparent, responsible and ethical manner. The benefits of an ethical based approach and data impact assessments include the ability to build trust and transparency with consumers, ultimately delivering long term benefits to both the consumer and the organisation.

Susan Bennett LLM(Hons), MBA

Principal of Sibenco Legal & Advisory and co-founder of Information Governance ANZ.

Susan is a lawyer and business advisor with twenty-five years of experience. She works closely with corporate and government clients to deliver tailored legal and risk management solutions that meet client needs and strategic objectives. If you would like assistance reviewing your current Information Governance ecosystem or developing an Ethical Framework or ethical based approach to data analytics within your organisation, please contact Susan on +61 2 8226 8682 or email susan.bennett@sibenco.com.