UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Gordon Institute
of Business Science
University of Pretoria

Big Data and Privacy
A Modernised Framework

Mandi Ainslie

23114772

A research project submitted to the Gordon Institute of Business Science, University of Pretoria, in partial fulfilment of the requirements for the degree of Master of Business Administration.

7 November 2016

ABSTRACT

Like the revolutions that preceded it, the Fourth Industrial Revolution has the potential to raise global income levels and improve the quality of life for populations around the world. Responding to global challenges, generating efficiencies, prediction improvement, democratisation access to information and empowering individuals are a few examples of the economic and social value created by personal information. However, this technological innovation, efficiency and productivity comes at a price -privacy. As a result, individuals are growingly concerned that companies and governments are not protecting data about them and that they are instead using it in ways not necessarily in their best interests.

The objective of this research is to investigate the validity and feasibility of a Personal Data Store (PDS) against the developed *a priori* framework.

Ten qualitative, semi-structured interviews using the long interview method were conducted with individuals identified as a subject matter expert (SMEs) in the Big Data analytics and the data privacy field.

The findings show that the guiding principles of transparency, control, trust and value, ensures the validity and feasibility of the PDS. Furthermore, user-centricity provides greater control within the Big Data continuum. However, as personal data should not be trusted in the hands of third-parties, identity management and security must be entrenched at a foundational level of the model. The remaining elements - selective disclosure, purpose and duration, signalling and data portability – is in fact value adding qualities that allows for the commodification of personal data.

In the age of the Internet of Things (IoT), organisations churn out increasing volumes of transactional data, capturing trillions of bytes of information about their customers, suppliers and operations. However, amplifying the rate of technological disruption with the failure to provide safe spaces where individuals can think free, divergent and creative thoughts will significantly diminish the progress organisations (and society) can enjoy.

KEYWORDS

Big Data; Privacy; Privacy Rights Management; Personal Data Store

DECLARATION

I declare that this research project is my own work. It is submitted in partial fulfilment of the requirements for the degree of Master of Business Administration at the Gordon Institute of Business Science, University of Pretoria. It has not been submitted before for any degree or examination in any other University. I further declare that I have obtained the necessary authorisation and consent to carry out this research.

Mandi Ainslie
7 November 2016

CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

GLOSSARY OF TERMS

User-centricity The foundational element for aligning stakeholders' interests and realising the vision of the personal data ecosystem. Also interchangeable with "The Individual"

Volunteered Data Information volunteered by individuals when they explicitly share information about themselves through electronic media, such as social network profile, credit card information and online purchases

Observed Data Data captured by recording activities of users; for example, internet browsing preferences, location data using cell phones and telephone usage behaviour

Inferred Data Data and insights derived from individuals based on the analysis of personal data.

Selective Disclosure The ability of customers to share their data selectively, without disclosing more personal data than they wish to

Purpose & Duration The purpose and duration of primary and secondary uses of a user's personal data

Signalling The means for individuals to express demand for goods or services in open markets, not tied to any single organisation

Identity Management The process to manage tasks such as the authentication and use of multiple identifiers while preventing correlation unless permitted by the user

Security Protective digital privacy measures that are applied to prevent unauthorised access to computers, databases and websites. This includes data protections against corruption.

Data Portability The ability to move all of the data from one provider to another using standard data formats and interface protocols

Accountability & Enforcement Accountability for protecting and securing personal data in accordance with the rights and permissions established by agreement and/or enforced by tagging mechanisms; and enforcement under self-regulatory guidelines and legal mandates, both back by comprehensive auditing.

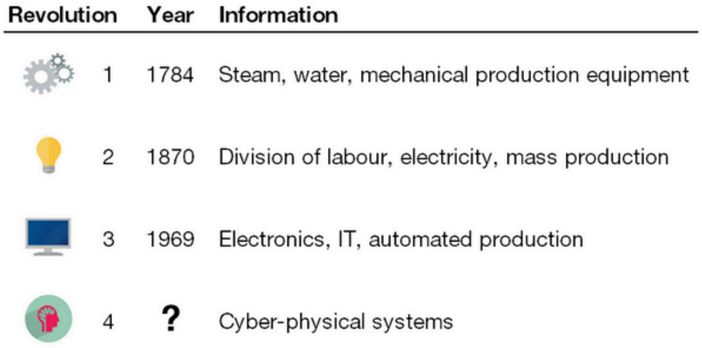
CHAPTER 1: INTRODUCTION TO RESEARCH PROBLEM

> *"I grew up with the understanding that the world I lived in was one where people enjoyed a sort of freedom to communicate with each other in privacy, without it being monitored, without it being measured or analysed or sort of judged by these shadowy figures or systems, any time they mention anything that travels across [the] public line."*
>
> – Edward Snowden (Greenwald, MacAskill & Poitras, 2013)

At the eve of its destruction, the Library of Alexandria was believed to house the sum of all human knowledge – projected at 400 000 scrolls or 200GB. Bearing in mind that a typical memory card for a camera or smart-phone is 32GB, there is enough information in the world today, to give every human being, three hundred and twenty times as much information as historians believe was stored in the entire Alexandria collection – estimated at 1200 Exabytes' worth of data (Cukier & Mayer-Schönberger, 2013).

Today, society stands on the brink of a technological revolution that will fundamentally alter the way we live, work and relate to one another. In its scale, scope and complexity, the transformation will be unlike anything humankind has experienced before (Schwab, 2016).

The first and second industrial revolutions utilised mechanised production and electric power to create mass production. The third industrial revolution introduced electronics and information technology, building the platform for the fourth revolution – characterised by a fusion of technologies that is blurring the lines between the physical, digital and biological spheres (Schwab, 2016).



Revolution	Year	Information
1	1784	Steam, water, mechanical production equipment
2	1870	Division of labour, electricity, mass production
3	1969	Electronics, IT, automated production
4	?	Cyber-physical systems

Figure 1-1: Navigating the next industrial revolution (Schwab, 2016)

The revolution that Big Data and digital transformation bring, is unlike the change in human communication reshaped by the internet. Cukier and Mayer-Schönberger (2013) suggest that it marks a transformation in how society processes information. Therefore, as society taps into ever more data to understand events and make decisions, we are likely to discover that many aspects of life are probabilistic, rather than certain.

This being said, data has become a torrent flowing into every area of the global economy. Rose and Kalapesi (2012) as well as Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh and Byers (2011) estimate that the Internet economy amounted to US$ 2.3 trillion in value in 2010, or 4.1% of total GDP, within the G20 group of nations. Larger than the economies of Brazil or Italy, the Internet's economic value is expected to nearly double by 2016 to US$ 4.2 trillion.

Similar to the revolutions that preceded it, the Fourth Industrial Revolution has the potential to raise global income levels and improve the quality of life for populations around the world (Schwab, 2016). Responding to global challenges, generating efficiencies, improvement in prediction, democratisation, access to information and empowering individuals are a few examples of the economic and social value created by personal information (Rose & Kalapesi, 2012).

However, this technological innovation, efficiency and productivity comes at a price: a loss of privacy.

In the age of the Internet of Things (IoT), organisations churn out increasing volumes of transactional data, capturing trillions of bytes of information about their customers, suppliers and operations. Millions of networked sensors are embedded in everything "smart", from our mobile phones and energy meters to automobiles and industrial machines that sense, create and communicate data (Tene & Polonetsky, 2012).

Additionally, the influx of Big Data has opened opportunities for a whole new class of professional gamers and manipulators, who take advantage of people using the power of statistics (O'Neil, 2016). As a result, individuals are growingly concerned that companies and governments are not protecting data about them and that the latter are, instead, using it in ways not necessarily in their best interests.

Essential to free and open societies, is freedom of thought. In the wake of widespread knowledge of governmental surveillance that was revealed by Edward Snowden in 2013, the conversation has primarily concerned the freedom of speech, but once organisations can access peoples' thoughts and emotions, a space is required that enables people to think freely, to engage in divergent and creative thinking. In a society where people fear having those thoughts, the likelihood of being able to enjoy progress is significantly diminished.

Stoycheff (2016) investigates the effects of subtle reminders of mass surveillance on subjects and illustrates the silencing effect of participants' dissenting opinions. The fact that the 'nothing to hide' individuals experience a significant chilling effect speaks to how online privacy is much bigger than the mere lawfulness of one's actions. It is about a fundamental human right to exercise control over one's self-presentation and image, in private, and now, in search histories and metadata.

These surveillance phenomena and statistical exploitation escalate when coupled with the inability of governments and organisations to safeguard individuals' personal data from unprecedented cyber-attacks. Citizens do not trust their governments with their personal data (Hall, 2016) and, as illustrated by the massive internet outage across the east coast of the United States on October 21, 2016, struggle to stay ahead of imminent cyber threats (Newman, 2016).

Even with legislation trying to strike a balance between protecting individuals and encouraging innovation and growth (Rose & Kalapesi, 2012) the ultimate collision looms. Tene and Polonetsky (2012) support this premise and argue that increasing privacy concerns could stir a regulatory backlash that would dampen the data economy and stifle innovation.

This raises the following issue: Can a balance be struck between the competing principles of Big Data and an individual's right to privacy?

The World Economic Forum report by Rose and Kalapesi (2012) argues that the explosive growth in the quantity and quality of personal data has created a significant opportunity to generate new forms of economic and social value. It is argued that just as tradable assets, such as water and oil, must flow to create value, so too must data. However, for data to flow well, it requires rules and frameworks for guidance through a plethora of privacy legislation.

The impact of the "right to be forgotten" (RTBF) on privacy and online information disclosure (Mangwanda, 2015) has limited its scope to the "right to erasure" aspect of the RTBF principle. Moreover, the focus of the study mentioned was primarily on the impact of the RTBF in the arena of social networking sites on an individual level and did not include the impact on Big Data strategy across industries.

Furthermore, Tene and Polonetsky (2012, p1) state that "in order to craft a balance between beneficial uses of data and individual privacy, policymakers must address some of the most fundamental concepts of privacy law, including the definition of 'personally identifiable information,' the role of individual control and the principles of data minimisation and purpose limitation". Likewise, Rubinstein (2013) postulates that Big Data challenges the Fair Information Practices, which form the basis of all modern privacy law.

Therefore, there is an academic and business necessity to investigate the overlapping requirements of the two completing principles.

The approach of this research was to review the available literature regarding Big Data, privacy as well as the legislative aspects, to identify if there were any references that could assist in answering the overarching research question.

An *a priori* framework was developed based on the available literature and, given that this is demonstrably a new area of research, it was decided that a qualitative, exploratory research method was the most appropriate. Semi-structured interviews with subject matter experts within the field of Big Data and privacy legislation were conducted. Following the data collection, the interview transcripts were analysed and coded for the themes identified. Furthermore, the findings have been presented and are interrogated against the available literature, followed by a discussion of the possibilities for future research and the implications for society at large.

It is hoped that this research will provide some insight for organisations and legislators in the search for a balance in the Big Data privacy conundrum. Quoting Schwab (2016, p 1): "We need to take responsibility at every level of society, from the individual and the personal to the institutional to the global to adapt to these technological challenges and changes which are redefining what it means to be human what it means to work, what it means to be completely embedded in this world".

CHAPTER 2: LITERATURE REVIEW

2.1 “Big Data” Concept and Discipline

The term “Big Data” is used to describe a wide range of concepts: from the technological ability to store, aggregate, and process data, to the cultural shift that is pervasively invading business and society, both of which are drowning in information overload (De Mauro, Greco & Grimaldi, 2015). Civilisation is moving towards a “Web of the world” - in which mobile communications, social technologies and sensors are connecting people, the Internet and the physical world into one interconnected network. Added to this, Gantz and Reinsel (2010) estimate that by 2020 the global volume of digital data will increase more than 40-fold. Therefore, although shrouded by much conceptual vagueness, Big Data is a trending buzzword in both academia and the industry.

De Mauro, Greco and Grimaldi (2015, p.103) propose a consensual definition where “Big Data represents Information assets characterised by High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value” by looking at the existing definitions of Big Data as well as at the main research topics associated with it:

- “Volume”, “Velocity” and “Variety”: describe the characteristics of Information involved
- Specific “Technology” and “Analytical Methods”: clarify the unique requirements strictly needed to make use of such Information
- Transformation into insights and consequent creation of economic “Value”: as the principal way Big Data is impacting companies and society.

In the case of transformation, value is further generalised to include “veracity” – the uncertainty of data. The four V’s of Big Data are illustrated in Figure 2.1 below as well as in Appendix 2.

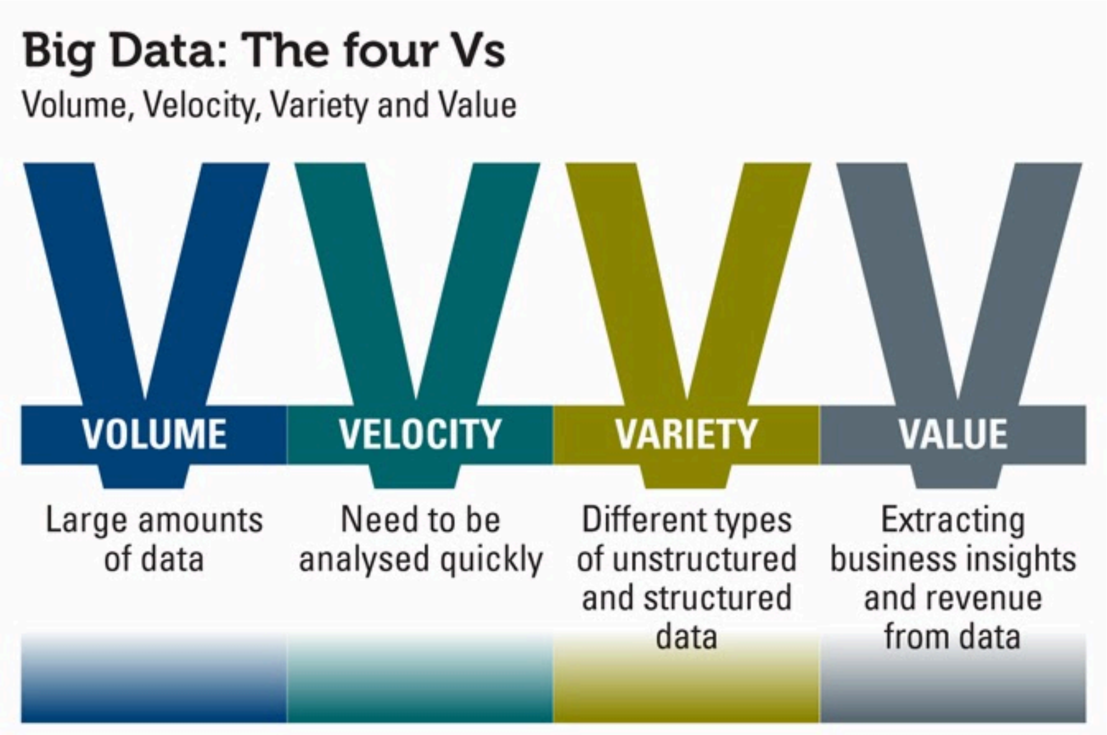

Big Data: The four Vs
Volume, Velocity, Variety and Value
VOLUME
Large amounts of data
VELOCITY
Need to be analysed quickly
VARIETY
Different types of unstructured and structured data
VALUE
Extracting business insights and revenue from data

Figure 2-1: The four V's of Big Data

(World Newsmedia Network, 2013)

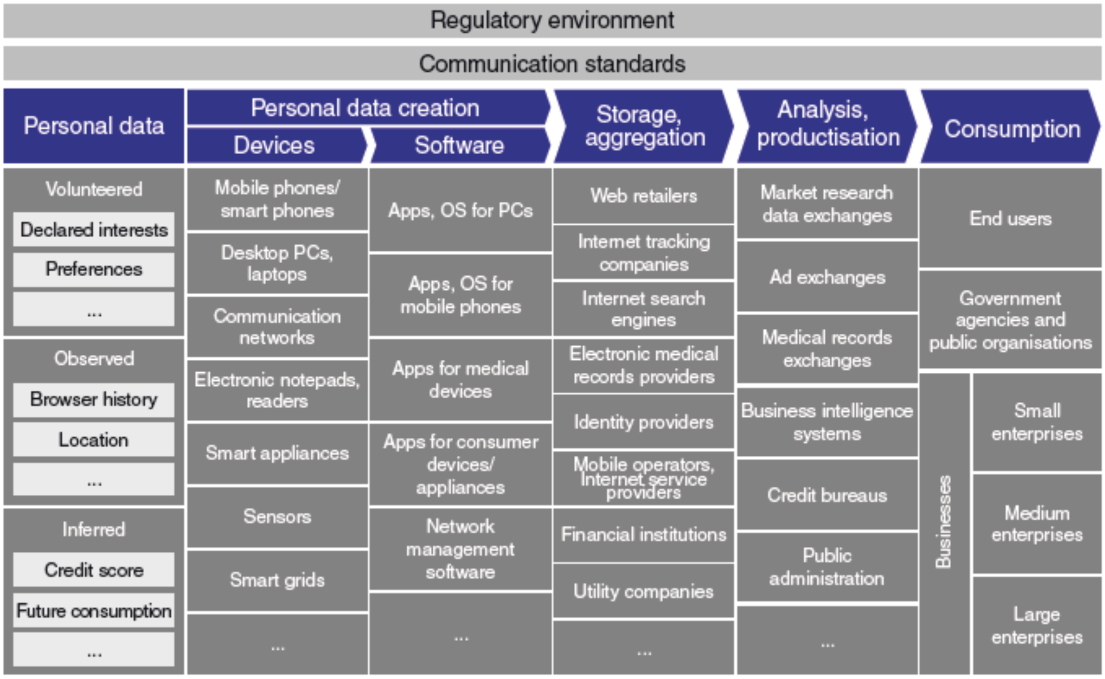
2.1.1 Information Globalisation

A fundamental reason for the Big Data phenomenon is the current extent to which information is being generated and made available. According to Gantz and Chute (2008) the International Data Corporation (IDC) reported that the overall created and copied data volume in the world was 1.8ZB (≈ 1021B), which increased by nearly nine times within five years. It is estimated that this figure will double every four years in the near future.

Digitisation (enabling analogue information to be transferred and stored in a more convenient digital format) and Datafication (organising digitised versions of analogue signals in order to generate insights that would have not been inferred while signals were in their original form) have become universal sensations due to the broad availability of devices that are both connected and provided with digital sensors. Digital sensors enable digitisation while connection permits data to be aggregated and, thus, permits datafication (De Mauro, Greco & Grimaldi, 2015).

Data is captured in a variety of ways. It can be volunteered by individuals when they explicitly share information about themselves through electronic media, such as social network profile, credit card information and online purchases, whereas observed data is captured by recording activities of users; for example, internet browsing preferences, location data using cell phones and telephone usage behaviour. Lastly, organisations also differentiate inferred data from individuals based on the analysis of personal data. For instance, credit scores can be calculated based on a number of factors relevant to an individual's financial history (World Economic Forum, 2011).

Each type of data, whether volunteered, observed or inferred, is created by multiple sources. The Figure 2.2 illustrates the complex personal data ecosystem from data creation to data consumption.



Regulatory environment

Communication standards

Personal data
Volunteered
Declared interests
Preferences
...
Observed
Browser history
Location
...
Inferred
Credit score
Future consumption
...

Personal data creation

Devices
Mobile phones/ smart phones
Desktop PCs, laptops
Communication networks
Electronic notepads, readers
Smart appliances
Sensors
Smart grids
...

Software
Apps, OS for PCs
Apps, OS for mobile phones
Apps for medical devices
Apps for consumer devices/ appliances
Network management software
...

Storage, aggregation
Web retailers
Internet tracking companies
Internet search engines
Electronic medical records providers
Identity providers
Mobile operators, Internet service providers
Financial institutions
Utility companies
...

Analysis, productisation
Market research data exchanges
Ad exchanges
Medical records exchanges
Business intelligence systems
Credit bureaus
Public administration
...

Consumption
End users
Government agencies and public organisations
Businesses
Small enterprises
Medium enterprises
Large enterprises

Figure 2-2: The personal data ecosystem

(World Economic Forum, 2011)

Evans (2011) estimated that between 2008 and 2009, the number of connected devices overtook the number of living people while, according to Gartner (2014), by 2020 there will be 26 billion devices on earth, more than 3 devices on average per person. The ubiquitous presence of a variety of objects (this includes mobile phones, sensors, Radio-Frequency Identification – RFID – tags, actuators) which are able to interact with each other and cooperate with their neighbours to reach common goals, goes under the name of the Internet of Things (IoT) (Estrin, Culler, Pister & Sukhatme, 2002) and (Atzori, Iera, & Morabito, 2010).

2.1.2 Underpinning technology, methods and skill

This eruption in data growth brings about opportunities for discovering new outcomes and trends, as well as an in-depth understanding of the hidden values. However, this exponential growth also provides for new challenges. Chen, Mao and Liu (2014) noted the following concerns:

- Integration of massive datasets from widely distributed data sources
- Storage, management and securing of huge heterogeneous datasets with moderate requirements for hardware and software infrastructure
- Effective "mining" at the various different levels during the analysis, modelling, visualisation and forecasting, so as to reveal its intrinsic property and improve the decision making, all whilst considering the heterogeneity, scalability, real-time, complexity and privacy of Big Data.

De Mauro, Greco and Grimaldi (2015) concur as another fundamental technological element is the ability to store a bigger quantity of data on smaller physical devices. Although Moore's (1965) law suggests that storing capacity increases over time in an exponential manner, the growing share of byte-hungry data types, such as images, sounds and videos, requires ongoing research and development to keep up the pace with information globalisation (Hilbert & Lopez, 2011).

The distributed nature of information requires a specific technological effort for transmitting big quantities of data and for monitoring the overall system performance using special benchmarking techniques (Xiong, Yu, Bei, Zhao, Zhang, Zou, Bai, Li, & Xu (2013). The open source framework most prominently associated with Big Data is Apache™ Hadoop®. The Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage (Taylor, 2010) Furthermore, machine-learning has capitalised on many domains such as science, business and government and is used to identify objects in images, transcribe speech into text, match news items, posts or products with users' interests and select relevant results of a search. Increasingly, these applications make use of a class of techniques known as deep learning (LeCun, Bengio, & Hinton, 2015).

Although the science is steadily maturing to cater for the technology footprint required to reap the value created by Big Data Analysis, Chen, Chiang and Storey (2012) remind us of the need for companies to invest in Business Intelligence and Analytics education. This includes the critical analytical and Information Technology (IT) skills, business and domain knowledge and communication skills required in a complex data-centric business environment.

De Mauro, Greco and Grimaldi (2015) in agreement with Buhl, Röglinger, Moser and Heidemann (2013) further suggest that the investment in analytical knowledge should be accompanied by a cultural change that would involve all employees and urge them to efficiently manage data properly and incorporate them into decision making processes.

2.1.3 The Big Data impact

The extent to which Big Data is impacting our society and our companies is often depicted through anecdotes and success stories of methods and technology implementations. When these stories are accompanied by proposals of new principles and methodological improvements they represent a valuable contribution to the creation of knowledge on the subject (De Mauro, Greco and Grimaldi, 2015).

2.1.3.1 Value Contribution of Big Data Analysis and Business Intelligence

Big Data has become a colossal industry. Research conducted at the Massachusetts Institute of Technology shows that companies that use “data-directed decision-making” enjoy a 5%–6% increase in productivity. There is a strong link between an effective data management strategy and financial performance. Dean, DiGrande, Field and Zwillenberg (2012) estimate that the Internet economy amounted to US$ 2.3 trillion in value in 2010, or 4.1% of total GDP, within the G20 group of nations. Larger than the economies of Brazil or Italy, the Internet’s economic value is expected to nearly double by 2016 to US$ 4.2 trillion.

Not only does Big Data and business intelligence have a significant impact on the micro and macro-economic landscape of the business environment, it also impacts various social aspects of society at large.

2.1.3.1.1 Healthcare

Big Data has unlimited potential for effectively storing, processing, querying, and analysing medical data. Healthcare organisations benefit greatly by developing actionable insights, organising their future vision, boosting up the outcomes and reducing time to value (TTV) responsiveness (Chen, Mao and Liu 2014).

Adverse Event Reporting Systems (AERS) affords an example of actionable insights created by analytical algorithms. For instance, AERS allows for early detection of adverse drug affects via signal detection algorithms. These systems identify statistically significant correlations between latent adverse effect signals from spontaneous reporting systems (Reshef, Reshef, Finucane, Grossman, McVean, Turnbaugh, Lander, Mitzenmacher & Sabeti, 2011).

In one such system, Microsoft Research examined de-identified Bing search engine logs, querying whether a higher proportion of users who searched for both "Paxil" and "Pravachol" also typed in words related to the "symptomatic footprint" (such as "headache" or "fatigue") than those who searched for just Paxil or Pravachol separately. The research hypothesis found support in that Big Data set. Users who searched Bing for the name of both drugs together were much likelier to search for diabetes-related side effects than users who searched for only one of the drugs (Tatonetti, Denny, Murphy, Fernald, Krishnan, Castro, Yue, Tsau, Kohane, Roden & 2011).

In another example, researchers in South Africa discovered a positive relationship between therapeutic vitamin B use and delaying progression to AIDS and death in HIV-positive patients (Kanter, Spencer, Steinberg, Soltysik, Yarnold & Graham, 1999). This was a critical finding at a time and in a region where therapies for people living with HIV are well beyond the financial means of most patients.

2.1.3.1.2 Geo-location Trend Analysis

"Always-On" mobile devices with multiple sensors, including cameras, microphones, movement sensors, GPS, and Wi-Fi capabilities, have revolutionised data collection and analysis (Tene & Polonetsky, 2012). Certain studies are currently analysing mobile phone communications to better understand the needs of citizens living in informal settlements within developing countries (Wesolowski & Eagle, 2010). These studies are exploring the various methods that could help predict food shortages using variables such as market prices, drought, migrations, previous regional production and seasonal variations (Okori & Obua, 2011).

2.1.3.1.3 Smart Grid Intelligence

Chen, Mao and Liu (2014) argue that Smart Grid is the next generation power grid consisting of traditional energy networks integrated with computers, communications and control for optimised generation, supply and consumption of electric energy. Smart Grid related Big Data is generated from various sources, such as:

- Power utilisation habits of users
- Synchronised real time measurement which is measured by Phasor measurement units (PMU) deployed national-wide
- Energy consumption data measured by smart meters in the Advanced Metering Infrastructure (AMI)
- Energy market pricing and bidding data and

- Management, control and maintenance data for various devices and equipment in the power generation, transmission and distribution networks (such as Circuit Breaker Monitors and transformers).

Pro-environment policymakers view the Smart Grid as key to providing better power quality and more efficient delivery of electricity to facilitate the move towards renewable energy (Tene and Polonetsky, 2012).

2.1.3.1.4 Traffic Management

Vehicles equipped with navigation systems containing embedded communication modules provide a range of telematics services to improve fuel-efficient driving and allow drivers to plan trips, taking into account the location of charging stations, or activate their air conditioner remotely. Planners benefit from the analysis of personal location data by way of decisions involving road and mass transit construction, mitigation of traffic congestion and the planning for high-density development (Duri, Elliott, Gruteser, Liu, Moskowitz, Perez, Singh & Tang, 2004; Ratti, Frenchman, Pulselli & Williams, 2006).

Such decisions not only reduce congestion but also control the emission of pollutants. At the same time, individual drivers benefit from smart routing based on real-time traffic information, including accident reports and information about scheduled road works and congested areas (Tene & Polonetsky, 2012).

2.1.3.1.5 Retail

Through Machine Learning and Big Data analysis, organisations are linking online activity to offline behaviour, in order to assess the effectiveness of online advertising campaigns, manage their supply chain as well as re-targeting in-store customers. Inventory Management Systems prescribe the flow of stock from warehouse to store to ensure that stores have the right amount of stock available to meet demand. Wal-Mart’s Retail Link system pioneered this process by enabling suppliers to see the exact number of their products on every shelf of every store at each precise moment in time, resulting in a significant decrease in their distribution costs (Tene, 2011).

Online behavioural advertising and personalisation applications depend on the knowledge of consumers’ personal preferences and behaviour typically distilled from volumes of granular information about them and stored in the form of consumer profiles (Adomavicius & Tuzhili, 2005). Consequently, personalised, targeted advertisements correlate directly with the amount of information collected from users. The more finely tailored the advertisement, the higher the conversion or “click through” rate and consequently, higher the revenues of advertisers, publishers, advertising intermediaries and ultimately, the supplier (Tene & Polonetsky, 2012). For example, Amazon utilises this Machine Learning and personalisation feedback loop quite successfully with the "*Customers Who Bought This Also Bought*" feature, prompting users to consider buying additional items selected by a collaborative filtering tool (Tene, 2011).

2.1.3.1.6 Payment Analysis and Fraud Prevention

Every year billions of dollars are lost worldwide due to credit card fraud, forcing organisations to continuously improve their fraud detection systems. (Bahnsen, Aouada, Stojanovic & Ottersten, 2016). As the use of credit and debit cards increases, so does fraud. According to the European Central Bank report (ECB, 2014), during 2012, the total level of fraud reached 1.33 billion Euros in the Single Euro Payments Area, which represents an increase of 14.8% compared to 2011. However, several detection systems based on Machine Learning techniques have been successfully used to counter this problem (Bhattacharyya, Jha, Tharakunnel & Westland, 2011).

2.1.3.1.7 Online Social Media Sharing

Since 2004, online social media, such as Internet forums, online communities, blogs, social networking services and social multimedia websites, have provided users with useful and easy opportunities to create, upload and share contents (Chen, Mao & Liu, 2014). Every day, individuals send or receive 196 billion e-mails, submit 500 million tweets and share 4.75 billion pieces of content on Facebook. Companies use personal data for a variety of purposes, amongst other things, to:

- reduce search costs for products via personalised and collaborative filtering of offerings
- lower transaction costs for themselves and consumers
- conduct risk analysis on a customer
- increase advertising returns through better targeting of advertisements. (Spiekermann, Acquisti, Bohme & Hui, 2015). However, personal data can also become a product in itself when it is linked with user-generated content.

While social networking sites share the basic purpose of online interaction and communication, specific goals and patterns of usage vary significantly across different services (Acquisti & Gross, 2006).

Disruptive innovations such as Airbnb and Uber have made the journey from an entrepreneurial start-up company to a multi-billion-dollar international corporation in less than five years, showing the real monetary impact that the online media market place has (Lashinsky, 2015; Konrad & Mac, 2014; Martin, 2016).

2.1.3.2 Adverse effects and risks of Big Data

Big Data can also impact society adversely.

2.1.3.2.1 Incremental effect

The accumulation of personal data has an incremental adverse effect on privacy. An example provided by Tene (2011) would occur where a researcher will draw entirely different conclusions from a string of online search queries consisting of the words "paris," "hilton" and "louvre" compared to one featuring "paris," "hilton" and "nicky." Adding thousands and thousands of search queries, a researcher can immediately sense how the data becomes ever more revealing. Moreover, once data – such as a clickstream or a cookie number – is linked to an identified individual, the pieces of information become difficult to disentangle. Illustrated by University of Texas researchers' Narayanan and Shmatikov's (2008) *Netflix Recommendation Experiment*, de-identified data was re-associated with identified individuals by cross-referencing a de-identified database with publicly available resources accessible online. Once any piece of data has been linked to a person's real identity, the anonymity of the virtual identity is removed.

Ohm (2010: p) describes this incremental effect as the "database of ruin", chewing away, byte by byte, on an individual's privacy until his or her profile is completely exposed. This effect has contributed to the concept known as "right to be forgotten" (Larson, 2013), (Mantelero, 2013) and (Mangwanda, 2015). Building on this concept, Bunn (2015) concludes that it is the right to not be indefinitely linked to information about one's past.

2.1.3.2.2 Automated Decision-making and Behaviour Modification

The influence of Machine Learning, predictive analytics that influence online behavioural advertising and personalisation applications, raises concerns around discrimination, self-determination and the narrowing of choice. The high customisation of advertisement and content that users will see influences decisions on the individual's credit, insurance and job prospects. Dividing individuals into pre-determined categories and automated decision-making compartmentalises society into pockets (or "echo chambers") of like-minded individuals, poses a risk to open society and democratic speech (Tene, 2011).

In an exposé published in *Politico Magazine*, Epstein (2015) and Rogers' (2015) fears were noted concerning Google's ability to influence voter behaviour in the 2016 US Elections. The order of search results, the ranking of positive or negative stories on the screen, can exert an enormous influence on the way individuals vote. In the scenarios where elections are close enough, the effect could be profound enough to change the outcome. In a study by Epstein and Robertson (2015) it was found that Google's search algorithm could easily shift the voting preferences of undecided voters by 20 percent or more – and up to 80 percent in some demographic groups – with virtually no one knowing they were being manipulated.

2.1.3.2.3 Predictive Analytics

Predictive analytics has various societal benefits and obvious positive implications for healthcare, specifically in the field of preventative care and early detection. However, predictive analytics also has unfavourable consequences.

Considering a recent New York Times publication, Duhigg (2012) found that retailers assign a "pregnancy score" to customers based purely on their purchasing habits. The article argues that although consumers' shopping habits are ingrained and notoriously hard to change, there are brief periods in a person's life when routines are in flux. An example of this is the birth of a child. Although birth records are usually public and couples are almost instantaneously bombarded with offers and incentives, the key for retailers is to know that a baby is on the way before other retailers do. Therefore, specially designed advertisements are required to target women in their second trimester. This is a sound business strategy promoting customer-centricity, personalisation and overall customer retention.

According to Duhigg (2012) Target's statisticians analysed historical buying records of women who had signed up for baby registries. The statisticians discovered latent patterns, such as women's preference for unscented lotion around the beginning of their second trimester or a tendency to buy supplements like calcium, magnesium and zinc within the first 20 weeks of a pregnancy. They were able to determine a set of products that, when grouped together, allowed Target to accurately predict a customer's pregnancy and due date.

Predictive analytics also has a useful application in law enforcement, national security, credit screening, insurance and employment. However, as noted by Tene and Polonetsky (2012) as well as Miller (2014), this raises an ethical dilemma where discrimination is prevalent in data profiling.

Although in conflict with the South African Constitution (1996), some predictive analytics perpetuate old prejudices. In further studies performed by Tene and Polonetsky (2012). it was found that the wealthy and well educated are more likely to be successful whilst the poor and underprivileged do not have the same analytical advantage. Additionally, by ignoring outliers predictive analysis becomes a self-fulfilling prophecy that accentuates social stratification and promotes inequality.

2.1.3.2.4 Lack of Access and Exclusion

Individuals are excluded from the benefits created from their own data in two ways: Firstly, individuals exchange personal data for free services. For example, as the online company knows the preferences of the transacting individual, goods and services are priced as close as possible to the individual's reservation price. Secondly, organisations are reluctant to share the valuable insights created by individual's personal data (Tene & Polonetsky, 2012).

Additionally, De Mauro, Greco and Grimaldi (2015) as well as Boyd and Crawford (2012) hypothesise that the split between information-rich and data-lacking companies could create a new digital divide that may slow down innovation in the sector. Specific policies will have to be promoted and data is likely to become a new dimension to consider within antitrust and non-competitive regulations.

The harvesting of large sets of personal data and the use of cutting edge analytics fuel growing privacy concerns. Protecting privacy will become harder as information is multiplied and shared ever more widely among multiple parties around the world. As more information regarding individuals' health, financials, location, electricity use and online activity percolates, concerns regarding profiling, tracking, discrimination, exclusion, government surveillance and loss of control arise (Daniel, 2006).

While one is appraising the benefits and risks attributed to Big Data, the question arises: why should society be concerned with the privacy conundrum?

2.2 The Matter of Privacy

Mangwanda (2015) reasons that the Internet is an information highway accessed by over 2.8 billion people (Euromonitor International, 2015) and that in this digital era, information has become easily accessible to any person who has a connection to the Internet. However, once this information is made available online, it can rarely be removed, thereby leaving one's footprints scattered across the Internet (Bergström, 2015). Building on Mangwanda (2015), a clear definition of privacy is required as well as insight to why privacy matters to society at large.

Westin (1968, p7) defines privacy as:

> *"...the claim of individuals, groups or institutions to determine for themselves when, how and to what extent information about them is communicated to others. Viewed in terms of the relation of the individual to social participation, privacy is the voluntary and temporary withdrawal of a person from the general society through physical or psychological means, either in a state of solitude or small group intimacy or, when among large groups, in a condition of anonymity or reserve."*

The privacy enigma has arisen to the level of a global debate, primarily due to the revelations of Edward Snowden (Greenwald, MacAskill & Poitras, 2013). The worldview that no real harm comes from mass surveillance is grounded in the premise that there are two kinds of people in the world: good and bad. Greenwald (2014) argues that the individuals who hold this view are actually engaged in a very extreme act of self-deprecation. This mind-set was clearly illustrated in a 2009 interview with CEO Eric Schmidt, then at Google, who, when asked about all the ways his company was causing invasions of privacy, said: "If you're doing something that you don't want other people to know, maybe you shouldn't be doing it in the first place" (*The Huffington Post*, 2009).

2.2.1 Liberation and democratisation

Literary illustrations of privacy infringement and its effects comprise a prominent theme in George Orwell's novel *1984*. It continues to spark concerns over loss of personal freedom that now extends to debates over privacy and the Internet. Cook (2002, p4) develops this ideal: "along with the disturbing surveillance prevalent throughout *1984*, privacy advocates have invoked the image of English, utilitarian philosopher Jeremy Bentham's Panopticon as another symbol of oppression." Furthermore, French philosopher and historian Michel Foucault expanded on Bentham's model prison by using the Panopticon as a device to illustrate the levels of power in society as well of conformism and compliance (Foucault, 1975; 2003; 2007).

Building on Orwell (1949), Dick (1956) explored the concept of a predictive crime model in the short story, "The Minority Report". Envisioning a world in which crime had been abolished, a compelling tale is told with a brilliant examination of the tensions between predictive ability, human agency and concepts of guilt and innocence.

It could be argued that Dick's (1956) vision is outlandish. However, in the United States, computer-generated risk scores are being used by some judges for sentencing and parole decisions. These reports match the individual's records against a mass of material on previous patterns of criminal behaviour and other demographic data. The premise is that such scores can predict the risks to society far more objectively than any judge. Properly designed algorithms should not be susceptible to conscious or unconscious bias (Thornhill, 2016). One glaring limitation of computer models are that they are only ever as good as the data they use. An investigation by *ProPublica* journalists Angwin, Larson, Mattu and Kirchner (2016) found that one software program that used historic data to predict future criminals systematically discriminated against black people.

Consequently, the argument that no privacy problem exists if a person has nothing to hide, is flawed. When the government engages in surveillance, many people believe that there is no threat to privacy unless the government uncovers unlawful activity, in which case a person has no legitimate justification to claim that it remains private (Solove, 2007). At a fundamental level, privacy does matter when considering the following:

- Limitation on power of governments and organisations
- Classical liberal respect for individuals
- Maintaining appropriate social boundaries – physical and informational
- Freedom of thought, speech, social and political views, commonly known as expressive behaviour.

The concept of *Liberté, égalité, fraternité*, is held up as a major tenet of a democratic society. Loss of privacy, then, accompanies the loss of democracy (Cook, 2002).

2.2.2 Privacy legislation

By focussing primarily on the European Union, United States and South Africa, current legislation models are constructed with the ultimate goal of permitting only legitimate possession of personal data. This goal is broken down into specific core principles:

- Data Quality – characterised in terms of purpose limitation, data minimisation, accuracy and completeness
- Consent
- Transparency
- Access and rectification

- Confidentiality
- Security.

Beyond these core principals, the European Union Data Protection Directive 95/46 EC (European Parliament European Commission, 1995) also seeks to ensure the free flow of personal data within the EU and addresses transfer of personal data to third countries, jurisdictional rules, administrative matters and enforcement. Moreover, the South African Protection of Personal Information Act (2013) (POPI) focuses on the constitutional right to privacy, balancing the interests and rights of data subjects as well as aiming to regulate the processing of data.

Although the legal premise of Privacy Legislation is sound, it has failed to keep pace with globalisation (Robinson, Graux, Botterman, & Valeri, 2009). With the relentless improvement and expansion of technological capabilities as well as the changing ways in which individuals create, share and use personal data, privacy legislation frameworks have been unsuccessful in meeting their goal.

2.2.2.1 Legislation follows practice

The main criticisms levied against privacy legislation are:

- Global information privacy legislation rests on the currently unstable category of Personally Identifiable Information (PII). (Schwartz & Solove, 2011)
- Reliance on concept of “Informed Choice” (Rubinstein, 2013)
- Technology evolves faster than law makers can regulate it (Moses, 2007).

2.2.2.1.1 Ambiguity of PII

Schwartz and Solove (2011) contend that information that falls within the PII category is protected, while information outside of it, is not.

- The anonymity myth is the incorrect assumption that as long as one does not explicitly do something under one’s actual name on the internet, there is safety from identification. However, the opposite is true. Due to the growth of static IP addresses, there is a basic level of built-in identifiability as soon as a computer connects to the Internet
- Re-identification of data. Technology increasingly enables marketers and others to combine various pieces of non-PII data sets to produce PII, or otherwise forge a link between some data and a specific person. In fact, the permanent de-identification of information is difficult because so much data about individuals exists in so many places and some of it is linked to specific identities

- As a result of technology innovation and information sharing practices, the line between PII and non-PI is not fixed
- PII context is blurred, as some data does not readily fit into one of the two categories.

The Protection of Personal Information Act (2013, p 8) is quite descriptive in that it defines personal information as "… information relating to an identifiable, living natural person and, where it is applicable, an identifiable, existing juristic person…"*.* On the other hand, the proposed European Commission's Data Protection Regulation expands the definition of personal data to include anything that "directly or indirectly" is "reasonably likely to be used" to identify a person, including an "identification number, location data and online identifier". The regulation, however, maintains that the principles of data protection should not apply to data rendered anonymous in such a way that the data subject is no longer identifiable (Kuner, 2012).

Furthermore, the ambiguity of PII appears evident when individuals engage in online transactions. Research illustrates that individuals are more willing to reveal PII on social media platforms such as Facebook when the receiving party shares the equivalent amount of information (Venkatanathan, Karapanos, Kostakos & Goncalves, 2013), which is a contradiction in terms of their heightened sense of privacy concern (Barnes, 2006) (Barnes, 2006). This pattern of behaviour obscures the lines of privacy as it increases an individual's exposure to exploitation if the receiving party has malicious intent regarding the disclosed information (Venkatanathan, Karapanos, Kostakos & Goncalves, 2013).

2.2.2.1.2 Informed Choice

The core principles approach outlined by Rubenstein (2013) is unreliable in the manner that it relies heavily on informed choice. Individuals neither read nor understand privacy policies, which rely on ambiguous language and are easily modified by firms (Cate, 2006) and (Winn, 2006). Consent is a hollow exercise.

2.2.2.1.3 Data Minimisation

Data minimisation is not enforced via re-design of software, hardware or business processes; globalisation and improvement of technological capabilities is changing the way individuals create, share and use personal information.

The Protection of Personal Information Act (2013) definition of "processing" is very wide and refers to almost any instance where personal information is handled. Additionally, it restricts the processing of data by limiting it to data to which a data-subject has consented or which has a valid justification such as the performance of a contract in compliance with law, or is in the interest of the data subject or a legitimate interest of a third party. Data must be collected directly from the data subject or be included, as part of public record, while analysis of data must have a specific and explicit purpose.

The “processing” concept results in various practical implementation concerns:

- Data mining and analysis may find it impossible to provide adequate notice for the simple reason that they do not (and cannot) know in advance what they may discover
- Users lack knowledge of potential correlations; they cannot knowingly consent to the use of their data for data mining or Big Data analytics
- Privacy laws apply solely to personal data, that is, to data relating to an identified or identifiable person. But there is no clarity on whether the core privacy principles apply to newly discovered knowledge derived from personal data, especially when that data has been anonymised or generalised by being transformed into group profiles; that is, profiles which apply to individuals as members of a reference group, even though a given individual may not actually exhibit the property in question (Chen, Mao & Liu, 2014).

Furthermore, from a theoretical point of view, the following is also questionable:

- Whether the personal data and non-personal data distinction remains viable in the scenario where there are potentially no limitations on the data scope
- Whether data anonymization remains effective in protecting users against tracking and profiling in the process of anonymization and
- Whether data minimisation – the idea that personal data processing must be restricted to the minimum amount necessary – can survive the onslaught of Big Data (Ohm, 2010).

It could be argued that the definition of PII should be expanded. However, as long ago as 1990, when data mining technologies were a great deal less sophisticated than they are today, 87% of the population of the US could already be uniquely identified by their gender, ZIP code, and date of birth (Sweeney, 2007). A disadvantage of a vastly expanded definition of PII is that the privacy framework would become all but unworkable. Consequently, concluding that a project raises privacy risks is not sufficient to discredit it. Privacy risks must be weighed against non-privacy rewards. And while numerous mechanisms exist to assess privacy risks, a balanced formula is lacking (Clarke, 2011) and (Golle, 2006).

2.2.3 Access, control, security and trust

In the current digital era, users engage in a privacy calculus conundrum, weighing the risks and benefits with each intention to disclose information online (Mangwanda, 2015). Much of the data gathered in computer databases is not particularly sensitive, such as one's race, birth date, gender, address or marital status. Many people do not care about concealing the hotels they stay at, the cars they own or rent, or the kind of beverages they drink. Often they do not take many steps to keep such information secret. In most instances, though not always, people's activities would not be inhibited if others knew this information.

Social contract theory hypothesises that individuals enter into relationships with organisations when information exchange occurs. The relationship can be explicit (legal) or implicit (social) in nature and is governed by certain principles, which include: defined ethical norms agreed on between the parties and informed consent by the parties with the ability to exit (Donaldson & Dunfee, 1994) (Mangwanda, 2015). Yuan (2012) defines social contracts as the universally understood obligations or social norms for the parties involved. Therefore, in assessing an institution's adherence to the principles of social contract theory, it can be concluded that ethical norms are agreed upon through the acceptance of privacy policies that have proved effective, if clearly worded and reduced in complexity (Capistrano & Chen, 2015).

However, there is some dispute regarding the social contract theory that focuses on the lack of clarity in the ownership and control of the disclosed data by individuals with other parties. According to De Wolf, Willaert and Pierson (2014), individuals become co-owners when they disclose personal information online, thus assuming that others have the same right as the disclosing individual to view and share such information.

2.3 Privacy Rights Management (PRM) and the Personal Data Store (PDS)

For business and society at large to continue reaping the benefits from Big Data and business analytics, an alternative framework and business model is required. Tene and Polonetsky (2012), Rubinstein (2013), Searls (2013) and Solove (2006) reason that a modernised framework and business intelligence strategy is important for two reasons:

1. Existing business models have proven time and again that privacy regulation is no match for them. Businesses inevitably collect and use more and more

personal data and while consumers receive many benefits in exchange, there is little doubt that businesses, not consumers, control the market in personal data with their own interests in mind

2. A modernised business model promises to shift collection and use of data from organisations to individuals. This will promote consumer empowerment and support a personal data ecosystem.

2.3.1 An *a priori* theoretical framework

As outlined by Searls (2012), supplemented by Tene and Polonetsky (2012) and formulated by Rubenstein (2013) a Personal Data Store or Service (PDS) business model is proposed that encompasses the eight primary elements to move individuals from an “Attention Economy” to “Intention Economy”. The vision and aim of this model (World Economic Forum, 2011) promises:

- Greater individual control over their personal data, digital identity and online privacy, as well as increased compensation or shared value for providing others with access to personal data
- Disparate silos of personal data held in corporations and government agencies will be more easily exchanged to increase utility and trust among people, private firms and the public sector
- Government’s need to maintain stability, security and individual rights will be met in a more flexible, holistic and adaptive manner.

The illustration below (Figure 2-3) provides an *a priori* preliminary framework to conceptualise the model proposed by Searls (2013).

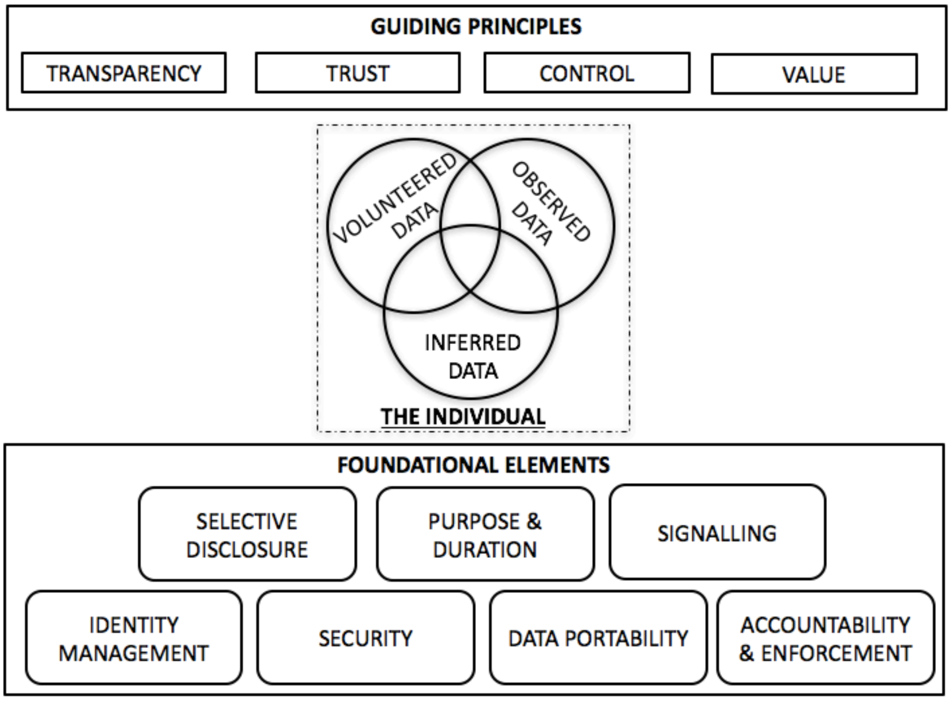


Figure 2-3: Preliminary *a priori* framework

Source: Researcher's own construction

The following section outlines the key framework elements in support of PRM and a PDS.

2.3.1.1 Individual as the centre of data collection, management and use

The foundation element for aligning stakeholders' interests and realising the vision of the personal data ecosystem is the concept of end User-centricity (World Economic Forum, 2011). This construct supports the designs for an intention economy, breaking away from the industrial-age model of the consumer – where relationships are captured, developed and owned. Figure 2.4 below illustrates User-centricity across diverse types of personal data.

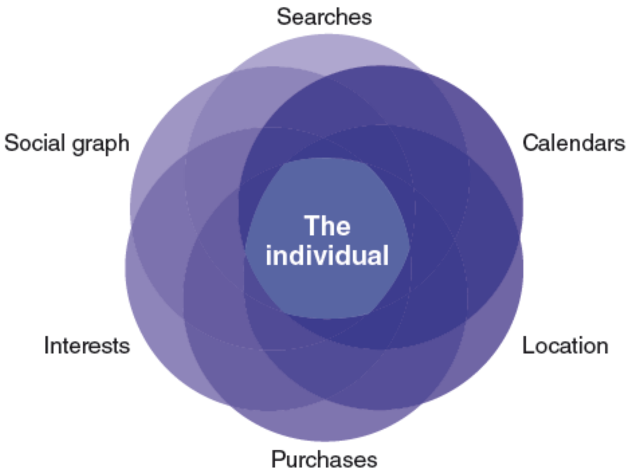


Figure 2-4: User-centricity across diverse types of personal data

(World Economic Forum, 2011)

The PRM and PDS would include all three types of personal data: volunteered, observed and inferred, and would be enabled via data warehousing and technology such as universal personal and mobile computing, to help individuals manage personal data as a personal asset (Rubenstein, 2013).

The right to access, update and rectify one's individual information remains distressingly underutilised (Tene, 2011). Few individuals are aware of their access rights and even fewer exercise them. A Eurobarometer Analytical Report (2008) found that across the European Union, just over a half of the citizens were aware of the right and far fewer had ever exercised it.

2.3.1.2 Selective disclosure

The concept of selective disclosure supports the ability of customers to share their data selectively, without disclosing more personal data than they wish to.

As recently as February 2016, Facebook (and other online media organisations) has started using an "Audience Network" platform to serve up ads to their users all over the web. Barrette (2016) notes that even although historically, users had the ability to opt out of this network that tracks behaviour across multiple websites and informs what advertising a user sees but not whether a user's actions on Facebook informs the ads they see. However, Facebook has gone a step further to allow users to opt out of both. Along with this change, it also serves ads to non-Facebook users, irrespective of whether a user has an account or not. Advertisements are based on one's online activity.

In contradiction to the Facebook “opt-out” model, selective disclosure requires the user to opt-in to online services and goods.

The proposition specifies that organisations subscribe to updates from specific fields within the individual’s Personal Data Store. To gain access they have to agree to the individual’s terms and conditions. The individual can choose which organisation he or she wishes to accept or reject as a subscriber. Once the subscription is in place, every time the individual changes the relevant field in her or his data store, the subscribing organisation is alerted to this fact. Searls (2013) suggests that on the premise that perfect data is provided during the opt-in action, operational costs associated with traditional data cleansing would not be necessary and will also remove the “guessing game” that data analytics play.

Tene and Polonetsky (2014) complement Searls (2013) and Rubinstein’s (2013) proposition of selective disclosure and similarly propose a “sharing the wealth strategy” premised on data controllers providing individuals with access to their data in a usable format and allowing them to take advantage of an application to analyse their own data and draw useful conclusions from it. This represents a fundamental shift in the management of personal data from a world where organisations gather, collect and use information about their customers for their own purposes, to one where individuals manage their own information for their own purpose and share some of this information with providers for joint benefits.

2.3.1.3 Purpose and duration

Building on Selective Disclosure, the third element advocates control over the purpose and duration of primary and secondary uses of a user’s personal data. This control may be achieved via an “owner data agreement” and/or by technical means such as Data Rights Management (DRM) of meta-data tagging.

Google has taken the first step towards “purpose and duration” by introducing a new data dashboard known as *My Activity*. This dashboard allows a user to see just about every single piece of data that Google has collected about him/her over the better part of the past two decades. This includes every website visited, every image viewed, every search term typed into the Google Search box (Purewal, 2016).

2.3.1.4 Signalling

Taking into account the proposition of “Purpose and Duration”, a PDS will enable “Signalling” - a means for individuals to express demand for goods or services in open markets, not tied to any single organisation.

The *raison d’être* signalling will promote higher quality advertisements as traditional targeting, capturing, acquiring, managing and locking a customer in would not be necessary. Customisation of products similarly empower vendors to better tailor product / service propositions to a customer and promote revenue returns.

2.3.1.5 Identity management

As part of the same process of interacting and transacting, organisations in both the private and public sector need to be confident that the person they are dealing with is who they say they are. Usually, this assurance is given by an agreed “gold standard” piece of identification such as a passport or bank account. Personal Data Stores can help streamline these identity assurance processes by linking verifications to such data (World Economic Forum, 2011).

Identity management within a Personal Data Store will manage tasks such as the authentication and use of multiple identifiers while preventing correlation unless permitted by the user.

2.3.1.6 Security

The personal data store is hypothesised on the premise that firstly; the access services would not have the right to use nor view the underlying data. Rather, according to Rubenstein (2013, p12) the system “would be structured to expose only those data elements authorised by privacy rules and policies, only authenticating authorised users, thereby allowing patients to opt in or out of these access services”.

Secondly, the data required by any access service may be gathered from various servers and aggregated, analysed, and presented in real time (similar to browsers assembling the elements of a Web page on a just-in-time basis). As a result, the ecosystem would remain decentralised. The third premise promotes to avoid the use of uniform patient identifiers. Lastly, tagged data elements would enable effective implementation of privacy rules and policies. Rubenstein (2013) in concurrence with Bygrave (2001) argues that this contradicts the largely empty ritual of privacy policies and the “all-or-nothing” choices typical of most Web sites today. This “tagging” system allows for fine-tuned individual privacy preferences.

2.3.1.7 Data portability

The proposition of "data portability" includes a user's ability to move all of the data from one provider to another using standard data formats and interface protocols. This flexibility allows for better competition and service delivery.

2.3.1.8 Accountability and enforcement

Rubinstein (2013) and Searls (2013) concur that accountability of all stakeholders as well as enforcement of protection and securing of personal data in accordance with the rights and permissions must be established by agreement and/or enforced by tagging mechanisms and enforcement under self-regulatory guidelines and legal mandates, both backed by comprehensive auditing.

Rubinstein (2013) concludes that it is noticeable how these eight elements align with the core European Union's data protection principles. Elements one to three (user-centricity, selective disclosure and purpose and duration) addresses purpose specification as well as collection and use limitations. Furthermore, user-centricity strongly supports data quality, while elements six and eight (security and accountability and enforcement) match up respectively with data security and accountability.

2.4 Conclusion

In conclusion, the overlapping requirements of the two competing principles – Big Data and personal data privacy – suggest that an alternative framework which includes business strategy as well as an individual's right to privacy as evidenced in Chapter 1, is required.

A conceptual preliminary *a priori* framework was presented in Figure 2.3 based on the various literature streams considered within in the Big Data privacy conundrum.

The said framework contends that the validity and technical feasibility of the hypothesised personal data store, all elements (user-centricity, selective disclosure, purpose and duration, signalling, identity management, security, data portability and accountability and enforcement) are required at a foundational level.

Research propositions based on the questions posed in Chapter 1 and the conceptual *a priori* framework and arguments presented in Chapter 2 are detailed in the following chapter.

CHAPTER 3: CONSTRUCT

3.1 Introduction

The review of the available theory during the literature review process has led the researcher to build an *a priori* framework to support the validity and technical feasibility of a Personal Data Store (PDS).

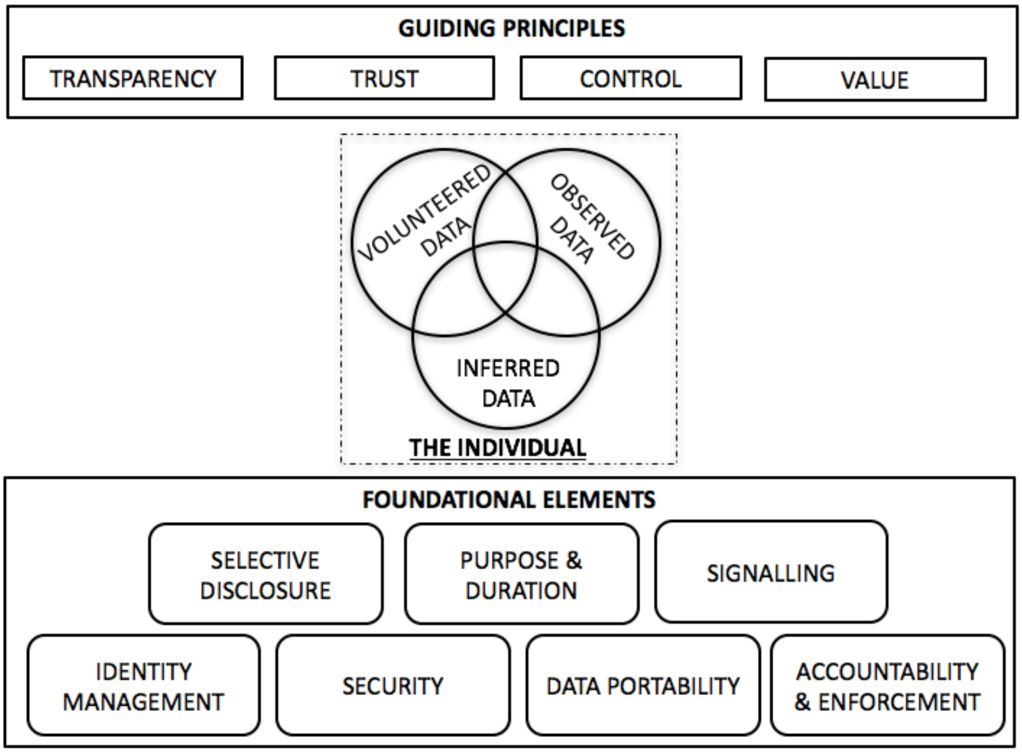


Figure 3-1: *A Priori* Personal Data Store (PDS) framework

Source: Researcher's own construction

3.2 Proposition 1: Guiding Principles

Based on the literature review it is proposed that a PDS would require the guiding principles of transparency, control, value and trust.

3.3 Proposition 2: User-centricity

From the literature review it is proposed that the central theme of a PDS is the construct of user-centricity.

3.4 Proposition 3: Foundational Elements

From the literature review, it is proposed that the validity and technical feasibility of the hypothesised PDS, all elements (selective disclosure, purpose and duration, signalling, identity management, security, data portability and accountability and enforcement) are required at a foundational level.

CHAPTER 4: PROPOSED RESEARCH METHODOLOGY AND DESIGN

4.1 Research Philosophy and Design

4.1.1 Method: qualitative, exploratory

Deliberating various research philosophies' multi-dimensional set of continua, the research methods reviewed suggested that exploratory studies should be considered where:

- A research problem requires exploration, when the aim is to discover new information about a subject or phenomenon (Saunders & Lewis, 2012, p.110); or
- A topic or population has been little explored and the researcher intends to listen to participants so as to gain an understanding based on what is heard (Creswell, 2014, p.29)

In contrast to quantitative research's positivist philosophy, qualitative research is more closely associated with an interpretive model (Denzin & Lincoln, 2011). Interpretivism advocates the necessity for a researcher to understand humans in our role as social actors. In the same way as we interpret our everyday social roles in accordance with the meaning we give to these roles, we also interpret the social roles of others in accordance with our own set of meanings. Furthermore, interpretivism is constructed from the intellectual traditions of phenomenology and symbolic interactionism. Saunders, Lewis and Thornhill (2012) therefore state that a study is interpretive when the researcher needs to make sense of the subject and socially constructed meaning expressed about the phenomenon being studied. Crucial to this philosophy is that the researcher is required to adopt an empathetic stance and perceive the world from the subject's point of view. Like quantitative research, qualitative research may also be utilised within realist and pragmatist philosophies within a multiple method research design.

As outlined in Chapter 2, Searls (2013), Tene and Polonetsky (2012) and Rubenstein's (2013) Big Data-privacy conundrum and resulting *a priori* framework has been extensively theorised. Conversely, the validity as well as the feasibility of the *a priori* framework, is a nascent concept as little research could be discovered at the time of the literature review. Therefore, it was decided that a qualitative, exploratory approach was best suited to this research, generating primary data, as no secondary data is available. By further considering the requirement of validity and feasibility of the *a priori* framework, it became clear that a deductive approach would be the most appropriate.

Saunders, Lewis and Thornhill's (2012) and Creswell's (2014) positions are adopted: that one of the main reasons to choose a qualitative approach exists when the study is exploratory. This usually means that very little has been written on a topic or the population being studied so that the researcher must listen to the respondents during interviews and create an understanding based on what is heard there. Furthermore, a qualitative method such as exploratory interviewing adds an element of credibility to the research study (Saunders & Lewis, 2012) since respondents are allowed an opportunity to provide a perception of their own reality, permitting a richness in the data analysis through the use of thick descriptions (Creswell & Miller, 2000).

4.1.2 Data collection and measurement instrument: Semi-structured, long interview

As the need for an exploratory, qualitative research method had been ascertained, a fitting means of data collection for this research was identified as non-standardised semi-structured interviews, more commonly referred to as qualitative research interviews (Cassell & Symon, 2004).

Individual, semi-structured, in-depth interviews, facilitated by an interview schedule with themes and key questions, were decided upon as the most appropriate because the researcher was unsure of the answers respondents would give (McCracken, 1988; Saunders & Lewis, 2012). In order to explore a topic or theme that the respondent raises, this method provides flexibility to omit certain questions that might be irrelevant to the participant's or respondent's particular context (Saunders & Lewis, 2012). Additionally, the long individual interview method provides the researcher with an opportunity to understand how a respondent thinks and views the world, whilst also allowing the latter to speak more freely and without being concerned about judgment and retribution from other members within the organisation. This offers contextual material for a deeper discussion (McCracken, 1988).

Following McCracken's (1988) long interview method, an interview schedule or discussion guide was prepared to assist during the interview. The purpose of the guide or schedule was to provide the researcher with a number of prompts that could be used, starting with a series of overview questions to set the scene. Once the respondent was more relaxed and forthcoming, framework validity prompts were used to sustain the conversation and affirm, contradict or add an element to the proposition. This interview schedule provided the interviewer with a proactive role; prompts were planned to focus the conversation in the direction necessary.

Saunders and Lewis (2012) suggest that it is good practice to review the discussion guide after each interview and, if necessary, adapt the questions. The order of the questions may vary, depending on the flow of the conversation, while additional questions may be required to explore the proposition as well as the objectives, given the nature of the event or particular organisation (Saunders, Lewis & Thornhill, 2012). This continuous review of results also identifies whether a researcher has reached the point of data saturation or not (Saunders & Lewis, 2012).

Following the first and second interview, the researcher deemed the interview guide as appropriate for the study: not needing any further changes to its structure and flow. The interview guide has been included in the Appendix 3. All interviews were recorded using a digital voice recorder, while all recordings have been digitally stored and submitted as part of the evidence of this study.

4.2 Population Universe

Introna and Pouloudi (1999) established that although the stakeholder concept has been extensively used and in a variety of contexts, it is within the area of strategic management that most development has occurred. Donaldson and Preston (1995) have captured this variability in a framework that distinguishes between a descriptive, an instrumental and a normative aspect to stakeholder theory.

- Descriptive consideration: describes the corporation (social unit) as a constellation of cooperative and competitive interests possessing intrinsic value
- Instrumental aspect: establishes a framework for examining the connections between the practice of stakeholder management and the achievement of various corporate (social unit) performance goals.
- Stakeholder theory: is fundamentally normative and involves acceptance of stakeholders as persons or groups with legitimate interests in procedural and/or substantive aspects of corporate and social activity, holding that the interests of all stakeholders are of intrinsic value.

Introna and Pouloudi (1999) therefore concluded that the conflict between the interests and values of different stakeholders may result in different perceptions of privacy and that the descriptive, instrumental and normative aspects could be viewed in "nested" circles.

The population (or universe of complete set of data) considered within this research includes all stakeholders within the Big Data continuum. However, considering the conclusions of Introna and Pouloudi (1999), the focus has been restricted to Big Data or business intelligence subject matter experts (SMEs) with:

- knowledge about the intricacies of privacy legislation; for example, legal practitioners in the field of data privacy
- corporate officers responsible for enterprise-wide governance and utilisation of information assets, such as Chief Data Officers
- subject matter experts in the interdisciplinary field of data science and statistical models
- individuals with a knowledge of their access and control rights in terms of their own individual data.

4.3 Sampling

4.3.1 Sampling method

In comparison to a census, Barnett (2002) argues that the use of sampling allows for a higher overall accuracy. The smaller the number of cases for which data collection is required, the more time a researcher is allowed to design and pilot the means of data collection as well as the level of detail collected.

Non-probabilistic sampling techniques – complemented by snowball and convenience sampling – were used to identify potential candidates to interview. As a full population list of SMEs with an understanding of Big Data was not available, a sampling frame was not applied as part of the sampling method. People known to the researcher were asked to provide references for possible respondents that would be regarded as a Big Data Principal and who consulted within the business intelligence industry or were practicing privacy attorneys.

The structure of the interview guide was used as an additional eligibility vetting tool by focusing the first set of questions to establish respondents' experience in the field. All respondents showed a clear and practical understanding of the subject matter.

4.3.2 Unit of analysis

The sample unit under study in this research is the individual respondents' perception of Big Data analytics within their organisation and their awareness of data privacy.

4.3.3 Sample size

Particular to qualitative research via semi structured interviews, the sample size is dependent on the research question and objectives (Patton 2002). Although the validity, understanding and insights gathered from data are more reliant on data analysis than sample size, Patton (2002) as well as Saunders and Lewis (2012) offer guidance as to the sample size to ensure that sufficient numbers of interviews are conducted. Saunders and Lewis (2012) maintain that most literature recommends establishing the number of interviews inductively or until data saturation has been reached. Furthermore, McCraken's (1988) observations are that eight interviews should be sufficient because qualitative research does not deal with issues of generalisability but of access, and that it is more important to work longer, with fewer people, than superficially with many.

Based on the time and resources available, the researcher was able to interview ten respondents over a period of two months. By the end of the tenth interview, nothing substantially new was being deduced; given the amount of data captured and the range of respondents interviewed, the researcher did not seek any further interviews.

The bulk of the interviews were held with respondents who had identified themselves as data scientists, privacy lawyers or enterprise architects within an organisation's management team. Table 4.1 provides the date of the interviews, the initials of the respondent and specialisation and role within the industry.

Table 4-1: Interview summary - ordered by date

Order	Respondent	Date	Industry	Specialisation and Role
1	GS	2016/07/25	Information Technology	Data Scientist
2	AT	2016/07/26	Information Technology	Data Scientist
3	MS	2016/07/27	Insurance	Security and Technology Infrastructure
4	RS	2016/07/27	IT Consulting	Enterprise Architect
5	SW	2016/08/01	Financial Services	Business Intelligence Practitioner
6	JJ	2016/08/02	Insurance	Business Intelligence Practitioner
7	TW	2016/08/10	Information Technology	Legal and Privacy
8	NB	2016/08/10	Legal	Privacy Legislation
9	DK	2016/08/24	Professional Services	Privacy Law
10	HV	2016/08/29	Information Technology	Enterprise Business Lead

4.4 Data Analysis

4.4.1 Analysis tool

Interview transcripts were analysed using the ATLAS.ti - computer-aided qualitative data analysis software (CAQDAS) program.

4.4.2 Transcription preparation

All interviews were completed and submitted to a transcription service. Completed transcriptions were validated against the audio recordings for accuracy. As recommended by Friese (2014) all transcripts were formatted in a similar manner to facilitate the use of analysis tools provided by ATLAS.ti such as auto-coding, code occurrence, code co-occurrence and the codes-primary documents table.

Respondents within the sample were anonymized and are only referenced by initials. Furthermore, transcripts have been sanitised in terms of any reference to names of respondents, company or group of companies that they work for, as well as any client names that might have been referred to. Personal information such as age, gender, education levels or race groups was not collected as this was not relevant within the research.

A detailed transcription process analysis is discussed in Chapter 5.

4.4.3 Method of analysis

Saunders, Lewis and Thornhill (2012) maintain that research commences from either a deductive or inductive approach. With regard to data analysis, the deductive method seeks to use existing theory to shape the adopted approach, whereas the creation of new theory aligns with the inductive approach.

The researcher elected to use the deductive analysis approach for the purposes of analysing the transcripts as existing theory was used to formulate the *a priori* framework and research objectives.

Bryman (1988, p81) criticises this approach, reasoning that: “The prior specification of a theory tends to be disfavoured because of the possibility of introducing a premature closure on the issues to be investigated, as well as the possibility of the theoretical constructs departing excessively from the views of participants in a social setting.” Contrary to Bryman’s view (1988), Saunders, Lewis and Thornhill (2012) find that commencing work from a theoretical perspective offers certain advantages. Linking research to an existing body of knowledge within the subject area provides an initial analytical framework. By the same token, Yinn (2009) advises a researcher to devise a theoretical or descriptive framework to identify the main variables, components and themes in the research project as well as the predicted or assumed relationships between them.

Building on Yinn (2009), Saunders, Lewis and Thornhill (2012) list three steps that should be applied to the deductive approach to search for patterns and themes within data. Firstly, developing meaningful codes to describe the data, then deciding on the appropriate unit of data to which categories can be attached; and lastly, attaching those categories to aforementioned unit. Creswell (2014) advocates that based on the theory reviewed, an initial coding table should be established, after which one should allow the codes to develop as additional information is discovered. This method suggests a mix of both a deductive and inductive approach to coding the data and should provide reasonable mitigation of Bryman's (1988) concerns.

Given the exploratory, qualitative nature of this research as well as the possibility that new themes and metrics would be found during the transcript analysis, the researcher adopted a mixed approach by developing an initial coding table for the purposes of deductive coding and building on it during the transcription analysis. The intent was that any codes that were created inductively would be identified separately from the deductive codes. A list of codes is available within Annexure 6. The details of the convention adopted by the researcher during the actual analysis have been detailed in Chapter 5.

4.5 Research Limitations

Lincoln and Guba (1985) have adapted the traditional canons of scientific inquiry for a more appropriate fit to interpretivist studies, and use alternative constructs such as credibility, transferability, dependability and confirmability. Considering these constructs, the research limitations were explored.

4.5.1 Researcher bias

Researcher bias includes any factor which induces bias in the researcher's recording of responses; because exploratory research is quite subjective it is influenced by her or his perspectives. It is therefore important for the researcher to acknowledge those potential biases, as their context will have an influence on how she or he interprets the findings of the research (Creswell, 2014, p. 188; Saunders & Lewis, 2012), and the researcher's culture may create as much "blindness as insight" (McCracken, 1988, p. 6).

Consequently, it must be recognised that the researcher has extensive experience working in Risk and Compliance as well as a broad understanding of the Information Technology industry. As a result, this may have biased some of the answers given by the respondent or may have placed too much emphasis on a particular theme.

4.5.2 Sampling bias

Yin (2009) notes that selecting new data collection units or interviewees as an offshoot of existing ones could be acceptable if the snowballing is purposeful, and not done out of convenience. To avoid the pitfall of convenience bias, the reason for selecting units or interviewees must be defined and critiqued prior to the interview. Yin (2009) recommends distinguishing between a purposive reason and a merely convenience one.

The use of snowball sampling resulted in a number of business partners that were closely involved in the researcher's multinational organisation and involved in some manner in the information technology industry. Three of the ten respondents were within the same multinational organisation as the researcher – while a further four were in close business partnership with the organisation. This may influence the transferability construct of the research and limit it to the said industry.

4.5.3 Respondent bias

All the respondents or participants showed a high level of comfort with the concepts and themes within the research. All but two of the respondents either actively use business intelligence models or lead teams that employ Big Data within the organisation. The remaining two respondents practiced within the current South African and European data privacy legislation framework.

While this may point to the fact that an understanding of technology as well as of legislative considerations related to privacy is an important aspect of an SME's understanding of business intelligence, no individuals or representatives of companies were interviewed that did not have a clear understanding of Big Data in practice; hence no data were available to provide a valid counterpoint. This touches on the transferability of the research findings.

4.6 Research Validity

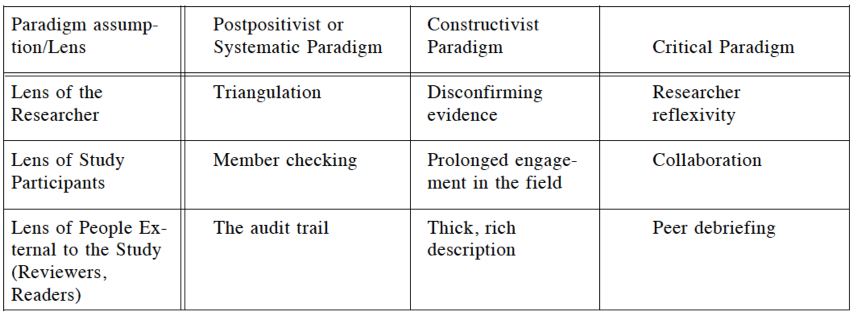
Various forms of validity have been identified to ensure the quality of research. Saunders and Lewis (2012) define validity of research as identifying whether the findings are really about what they seem to be, and urge researchers to establish construct- , internal and external validity.

Construct validity is concerned with the extent to which research measures actually measure the intent, whereas internal validity is established when a causal relationship is demonstrated between variables. These two concepts are associated with both positivist and quantitative research and can be applied to causal or explanatory studies. However, these validity measures are not suitable for exploratory or purely descriptive studies and were therefore not considered within this research (Saunders, Lewis & Thornhill, 2012).

Building on the overall research quality, Schwandt (1997) defines validity as establishing how accurately the account represents participants' realities of the social phenomena and whether it is credible. Creswell & Miller (2000) provide a two-dimensional lens framework to identify the appropriate validity procedures for qualitative research.

Table 4-2 presents the validity procedures from a qualitative lens and paradigm assumption.

Table 4-2: Validity procedures within qualitative lens and paradigm assumptions (Creswell and Miller, 2000)



Paradigm assumption/Lens	Postpositivist or Systematic Paradigm	Constructivist Paradigm	Critical Paradigm
Lens of the Researcher	Triangulation	Disconfirming evidence	Researcher reflexivity
Lens of Study Participants	Member checking	Prolonged engagement in the field	Collaboration
Lens of People External to the Study (Reviewers, Readers)	The audit trail	Thick, rich description	Peer debriefing

From the nine lenses provided, the researcher deemed four appropriate.

Researcher reflexivity is concerned with a researcher's assumptions, beliefs and biases and how these influence collection and interpretation of the data gathered during the research process (Creswell & Miller, 2000). This is discussed under researcher bias in section 4.5.1.

Disconfirming evidence (a procedure closely related to triangulation) involves a process whereby the researcher first identifies the preliminary themes and then searches for data that are inconsistent with and disconfirm these themes (Miles and Huberman, 1994; Creswell & Miller, 2000). The research accommodated this validity by adopting a mixed approach: Firstly, the initial coding scheme was developed deductively from the theoretical descriptive framework to identify the main variables, components and themes. It was subsequently expanded via an inductive approach and then refined over a number of additional coding passes. Codes identified after the initial deductive approach have been separately identified with a suffix of "*" on the code. Lastly, a final coding pass was done where any quotes that disconfirm a theme or topic were identified for inclusion in Chapter 5 where these have been included under the relevant sections, along with the evidence from the transcripts.

According to Denzin (1989), thick descriptions are deep, dense, detailed accounts whereas thin descriptions, by contrast, lack detail, and simply report facts. Thick, rich descriptions require additional information regarding the setting, the participants and the themes of the study in rich detail (Creswell & Miller, 2000). Chapter 5 details the setting, participants and themes of the study and in addition includes further contextual information. The quotes provided have also been left as close to the original as document space would allow while any contraction has been shown using ellipses (…).

The audit trail approach suggested by Creswell and Miller (2000) has been applied by providing a detailed account of the document management and analysis process that was followed during the research in Chapter 5 as well as the interview schedule utilised during interviews. All the original recordings, copies of the transcripts as well as database units and analysis from ATLAS.ti have been included in the evidentiary documentation. The "audit trail" lens establishes validity by providing clear documentation of all research decisions and activities (Creswell & Miller, 2000, p. 128).

4.7 Ethical Considerations

Ethical considerations relate to the standards of behaviour that guide the conduct in relation to the rights of those who are the subject of the research performed or who are affected by the research itself. Saunders, Lewis and Thornhill (2012) emphasise the appropriateness or acceptability of a researcher's conduct as it is influenced by broader norms of social behaviour.

With this in mind, a consent form was given to each participant at the start of the interview for the respondents who were interviewed in person, or emailed ahead of the interview in the case of the telephonic/Skype interviews. This allowed the respondent time to review the document before the interview began. Each consent form has been scanned and included as part of the evidentiary documentation for this research project. Please see Appendix 4 for the template of the consent form that was used. Furthermore, no information regarding the age, race or gender of the respondents was recorded as this was not a requirement in the literature reviewed during this research. All interview transcripts have also been anonymised to maintain the respondents' confidentiality. Respondents are only referred to by their initials along with any contextual information outlined.

CHAPTER 5: RESULTS

5.1 Introduction

The interviews performed as part of this research project have provided some valuable insight into how individuals view personal privacy as well as the validity and feasibility of a PDS and the influence it would have on the privacy of individuals.

This chapter builds on the research methodology outlined in Chapter 4, offering a summary of the interviews undertaken and contextual details of the respondents as well as a discussion of the processes followed by the researcher to ensure the accuracy, completeness and validity of the data collection and transcription. The section is followed by an exploratory discussion of the interviews in the context of the *a priori* research framework proposed in Chapter 2.

5.2 Summary of Interviews Conducted and Interview Method

The researcher initially intended to conduct a minimum of eight interviews or until the point of data saturation was reached (McCracken, 1988; Saunders & Lewis, 2012). A total of ten interviews were conducted with individuals identified as a subject matter expert (SMEs) in the Big Data analytics and data privacy field.

By the end of the tenth interview, nothing significantly new was being heard. Given the amount of data captured and range of respondents interviewed, the researcher did not undertake any further interviews. Table 5.1 provides information regarding the respondents who were interviewed.

Table 5-1: Summary of respondents and interview statistics

Respondent	Industry	Specialisation and Role	Length	Word count
Respondent_GS	Information Technology	Data Scientist	00:52:11	8109
Respondent_AT	Information Technology	Data Scientist	00:52:52	7946
Respondent_MS	Insurance	Security and Technology Infrastructure	00:44:32	7876
Respondent_RS	IT Consulting	Enterprise Architect	00:45:54	6512
Respondent_SW	Financial Services	Business Intelligence Practitioner	00:51:60	6853
Respondent_JJ	Insurance	Business Intelligence Practitioner	01:31:13	15066

Respondent	Industry	Specialisation and Role	Length	Word count
Respondent_TW	Information Technology	In-house Council	00:42:20	7799
Respondent_NB	Legal	Privacy Legislation	00:56:15	6928
Respondent_DK	Professional Services	Privacy Legislation	00:30:29	4863
Respondent_HV	Information Technology	Enterprise Business Lead	00:33:12	5286
		Average	00:49:53	7724
		Total	07:28:55	77238

The interviews were conducted over a period of a month and followed a semi-structured interview format as recommended by McCracken (1988).

A total of 448 minutes (just under 7.5 hours) of audio recordings were made and the resulting transcripts totalled 77238 words. The average interview length was 49 minutes long with an average transcript length of 7724 words.

Of the set of interviews conducted, nine were facilitated via an in-person meeting in a private meeting room. Logistical constraints saw one interview conducted via Skype. As no notes were taken during the interview sessions, all interviews were recorded using a digital voice recorder; the recordings were downloaded from the device and backed up to the Cloud before leaving the interview location.

Qualitative research literature provided several guidelines on how to best ask questions during an interview (McCracken, 1988; Meyers, 2013; Saunders & Lewis, 2012). Following the first interview, the researcher reflected on the appropriateness of the interview schedule, the articulated questions and the consistency of answers provided. As the conversation flow was suitable and rich answers were provided during the interview, the schedule was deemed appropriate and no further changes were made. Please see Appendix 3 for the interview schedule developed.

The respondents were overwhelmingly candid and forthcoming with answers: their outspokenness might be attributed to their level of understanding and knowledge within the subject field. Moreover, respondents related technical or more difficult concepts to real world scenarios. For example, respondent six (JJ) related his impressions to how he deals with his staff and even his mother on a daily basis. This added credibility to his understanding of and insights into the concepts discussed.

5.3 Interview Transcription and Verification

All interview recordings were sent to a transcription service with an enclosed confidentiality agreement. Please see Appendix 5 for copy of agreement.

Transcription accuracy was verified by the researcher against the original recordings and any mistakes (spelling, inaudibility or otherwise) were corrected. All transcripts were reformatted for consistency in the form of font size and type and line spacing.

5.4 Transcript Coding and Analysis in ATLAS.ti

As outlined within the research methodology (Chapter 4), the transcribed interviews were analysed using ATLAS.ti.

5.4.1 Transcript preparation and management

Following the recommendations of Friese (2014), all verified transcripts were imported into ATLAS.ti. and named according to the following convention:
"Initials_Specialisation_Interview Date"

- Initials: unique initials of the respondent – used during the transcript analysis to identify quotes and statements;
- _Specialisation: SME focus within the Big Data context;
- _Interview Date: the date of the interview

Each paragraph within the transcripts was identified with either MA (the researcher's initials) or the initials of the respondent, for example JJ or MS.

5.4.2 Transcript coding and code development

Given the exploratory, qualitative nature of this research as well as the possibility that new themes and metrics would be found during the transcript analysis, the researcher adopted a mixed approach by developing an initial coding table for the purposes of deductive coding and building on it during the transcription analysis. The intent was that any codes that were inductively created would be identified separately from the deductive codes. The list of codes is available in Appendix 6.

Furthermore, Friese (2014) advocates that a code swam should be avoided and that different aspects should not be lumped under one code name. Several layers within coding ensure that content and attributes are segregated to provide for rich analysis. The researcher customised this approach and developed an initial coding scheme deductively based on the literature reviewed. Inductive codes were then developed where new themes where discussed. Table 5.2 outlines the deductive and inductive code development. Please see Appendix 6 for a full list of developed codes.

Table 5-2: Deductive and inductive code development

Prefix / Suffix	Description
Code Number "C(Number)"	All codes are distinguishable with a C(Number)
"_Descriptive Category"	Codes prefixed by a [category]: are either used to identify sets of quotes, such as evidence of the role of the respondent in the organisation, or the maturity of the respondent. This forms part of the transcript naming convention mentioned in section 5.4.1. Codes prefixed by a [descriptor]: attempt to identify the sentiment or effect of the statement made by the respondent on the theme they are discussing. Examples are importance, or whether they feel it is a positive or negative effect.
Asterisk "*"	*As a suffix*: "inductive codes" developed during coding analysis. These codes were created after the initial coding table was defined and may be consolidated under a primary code or be given their own category as the themes were identified.

The auto-coding function within ATLAS.ti was not utilised as the researcher opted to apply the deductive codes manually to each individual's statement. The transcripts were then fully coded using the following two steps:

- An initial deductive coding pass of all transcripts was performed to identify and code sections of the conversational text that are strongly related to the concepts and themes identified in the literature reviewed
- A second coding pass was performed – inductively – to code themes and concepts that were not part of the initial deductive code selection. Inductive codes have been highlighted with an asterisk "*".

5.5 Details and Contextual Information of the Respondents Interviewed

From a contextual perspective, the industry and specialisation details of the respondents interviewed have been provided to offer a sense of who the respondents were, what role they played in their organisations and of their context.

Table 5-3: Contextual respondent information

Industry	Specialisation and Role	Respondent
Information Technology	Data Scientist	Respondent_GS
Information Technology	Data Scientist	Respondent_AT
Information Technology	In-house Counsel	Respondent_TW
Information Technology	Enterprise Business Lead	Respondent_HV
Information Technology	Enterprise Architect	Respondent_RS
Insurance	Security and Technology Infrastructure	Respondent_MS
Insurance	Business Intelligence Practitioner	Respondent_JJ
Legal	Privacy Legislation	Respondent_NB
Professional Services	Privacy Legislation	Respondent_DK
Financial Services	Business Intelligence Practitioner	Respondent_SW

The researcher found that the two respondents marked as Business Intelligence Practitioners, tended to express the same view on certain questions and concepts posed. This might be attributed to the following:

- Similarities between the Financial Services and Financial Industries
- Business intelligence speculation has a lengthy history with set, established, frameworks.

Likewise, Respondent_NB and Respondent_DK – both practicing attorneys – tended to have more confidence in enforcing legislation and regulatory institutions.

5.6 Deductive and Inductive Analysis Results

5.6.1 Understanding of Big Data

At the start of each interview, the respondent was asked to describe the contextual understanding of Big Data and business intelligence. Respondents were very forthcoming and displayed an appropriate understanding of the various intricacies and concepts.

According to Respondent_MS,

> *"… intelligent BI, intelligent Analytics is certainly what the future holds. I think that the days where human bodies and fingers sit and eye data and look for trends and try and analyse and then support a business process almost – if I can use the expression "manually" are long gone." (Respondent_MS)*

Respondent_RS elaborated:

> *"… data is like the new oil, it's the new wealth… the new power and I'm talking about Big Data, small data, structured data, unstructured data. Data and the insights that they yield will be responsible for the largest transfer of wealth since the Industrial Revolution. Companies that will understand how to actually leverage that data for competitive advantage… are the companies that will survive. And not just from a company perspective, elections will be won by understanding data trends and insights. If you look at the partnerships election candidates are forming with search houses, search engines that's evidence right there, the same thing with sports so data in itself if you want to know what's next the capitalization of data in one from or the other whether that be predictive analytics, prescriptive analytics, Machine Learning or artificial intelligence, all boils down from data." (Respondent_RS)*

However, Respondent_JJ specified that the collection of static data alone without analytical methods for its transformation into "value" did not constitute Big Data in its truest form:

> *" …from an organisation point of view… Big Data is a point where there is more data than what you have infrastructure for… a lot of times people miss the idea of Big Data and they are still applying all the dashboard reporting thinking methodologies onto a brand new concept called Big Data. We had a similar issue when people moved away from reports to infographics, I don't know if you have heard of that infographic rise, and everybody created infographic but they just looked like dashboards with pretty colours, but that doesn't make it infographic, and I think a lot of times that what we are doing now is we are taking Big Data and we are asking small data questions." (Respondent_JJ)*

This view was supported by Respondent_GS and Respondent_TW:

> *"… the challenge for most organisations now is how to embed these competencies within processes allowing for real time analytics at a point when it's needed so that it can be translated into actionable management actions almost instantaneously… Big Data has been around for a long time too and there is a difference between analytics which is… forward looking whereas business intelligence is more… backward looking." (Respondent_GS)*

> *"Just collecting data, which is just personal information so I can go do some marketing is not really Big Data… there is that absolute distinction." (Respondent_TW)*

Respondent_TW further explained the need for traditional Big Data:

> *"So I think your ability to break things down and categorise it and release it… is where Big Data is going. If it was pure traditional Big Data, we want to do massive cancer research, we want to genome genetic research, sure no problem in that space because that is very traditional Big Data… collect and using personal information to arrive at a Big Data benefit." (Respondent_TW)*

Conclusion:

Respondents are in agreement that Big Data represents information assets characterised by high volume, velocity and variety that require technological and analytical intervention (or rather veracity) for its transformation into value.

5.6.2 Benefits of Big Data

Several benefits of Big Data and business intelligence were noted; specifically, the impact on health care, geo location, Smart Grid technologies, traffic management retail, payment analysis and fraud analytics.

Respondent_NB and Respondent_GS agreed with these constructive attributes of Big Data:

> *"I agree Big Data analytics is a good thing, purely because the power of information that it gives you and if you run the right analytics on that information that you have it allows you make all these predictions of*

> *things not thought possible in the past, the power of that makes it a huge benefit." (Respondent_NB)*

> *"… those are very good examples… big companies like FedEx and Google are the obvious ones, they've been using data… since they've started in the mid-2000s using your data and everything that you do as a client… as a consumer of their free services… Amazon is another big one, these are all good examples of how companies… live and die on their data." (Respondent_GS)*

Respondent_AT expressed the view that Machine Learning is contributing to this positive impact:

> *"… Business Intelligence, Machine Learning, Big Data and all of these things like traffic management systems are perfect examples of Big Data becoming a real world example of how to make things better…" (Respondent_AT)*

The concept that Big Data "understood" an individual's context resonated with Respondent_RS and Respondent_HV, as the value that Big Data would bring to their world was immeasurable:

> *"It's much more valuable for me to have my phone understand that I like Italian food… for example because I visit Hello Tomato at least three times a week, and then suggest Italian recipes through apps to me and maybe connect with some courses and stuff that they are highlighting for the next upcoming week… my device understands my context…" (Respondent_RS)*

> *"So you're busy texting and driving and there's this thing on your screen, flashing "STOP DOING IT!"… that could be a useful… taking some of the behavioural data that they have of you and…prevent you from accessing anything but just take a call for example…" (Respondent_HV)*

Respondent_MS explained that trend analysis within several data fields had the power to highlight "golden threads" to identify and eliminate security incidents and events.

> *"From a security perspective, it's actually enabling us in other ways. From a Big Data perspective what's really cool is that were starting to*

> *get an idea of what normal baseline behaviour looks like… Big Data is really giving us some flexibility in terms of trying to identify anomalous type behaviour…*
>
> *…from a business perspective were getting advanced intelligence in terms of like what do members perhaps spend their monthly miles on… That kind of stuff but from my side there's obviously that broader benefit that we see just from the anomaly type and you kind of touched on it when you talked about Machine Learning…" (Respondent_MS)*

Respondent_MS further elaborated:

> *"…key point comes in terms of the data projection… making sure only authorised people have access to that data, then from a security perspective we have our own Big Data and analytic type systems… we're bringing in various data sources across the network from a security perspective and we're looking for specific incidents and golden threads …" (Respondent_MS)*

This sentiment was again echoed by Respondent_RS, Respondent_JJ and Respondent_HV:

> *"You see the thing is data isn't valuable, it's the insights… and trends that the data yields…" (Respondent_RS)*
>
> *"… we have only started to scratch the surface of what it can offer and I think we are not exploiting it nearly enough because my data is a natural resource like gold and oil and everything; there's work involved to extract it, to get the value from it and I just don't think we are doing nearly enough with the vast amount of data we have." (Respondent_JJ)*
>
> *"What do I think about Big Data? I think we have only started to scratch the surface of what it can offer and I think we are not exploiting it nearly enough… my data is a natural resource like gold and oil… there's work involved to extract it, to get the value from it and I just don't think we are doing nearly enough with the vast amount of data we have." (Respondent_HV)*

Conclusion:

Not only do Big Data and business intelligence have a significant impact on the micro and macro-economic landscape of the business environment, they also progressively impact various aspects of society at large.

5.6.3 Personal privacy

Respondents overwhelmingly agreed with the benefits of Big Data and business intelligence (for both the organisation as well as for the individual), but were equally concerned with the adverse effects, such as civil liberty intrusions as well as infringement on their personal privacy.

Respondent_AT expressed apprehension concerning the progress of artificial intelligence and the impact on one's / their civil liberties:

> *"… then there's a conversation to be had because now let's say your civil liberties are being affected and you're being affected as a personal individual and the technology is possibly getting in the way of your own progress as an individual, but for the most part I don't think these kind of technologies really are at the point yet where are so intrusive of our lives there's really an ethical conversation to be had. When we start going down the line of Artificial Intelligence as an example that is where the lines start getting blurred because now you're giving machines the capability to think at the same general level as human mental capacity…" (Respondent_AT)*

Respondent_TW echoed this concern:

> *"As I said I just think that on the AI side and the Big Data side of things we just need to be very, very careful and we need to be little bit wiser in how we approach these things. Sometimes going a little bit slow around things we don't understand. The future consequences of this probably what's called for. So hopefully the law will think ahead." (Respondent_TW)*

Each respondent articulated quite a progressive view on how they viewed their own online and mobile app privacy.

Respondent_MS and Respondent_DK took their own privacy very seriously:

> *"I'm a bit of a privacy zealot, I would say, maybe zealot is the wrong word better than "nut" right? …I use communication* [that] *is encrypted end-to-end, so I use signal to talk to my buddies and I don't use open sourced platforms like Google Talk or WhatsApp. Well I use WhatsApp now because its end-to-end encrypted, but typically I'm quite keen to make sure that my private information stays private." (Respondent_MS)*

> *"…personally I'll apply the rules, I'll make sure the privacy settings are in place on social media sites and things like that. I'm not overly fanatical about it so I'm not one of those that will remove myself off the grid because I'm concerned about my data being sold but I'm very conscious of it, personally because of the industry that I was in." (Respondent_DK)*

Whereas Respondent_JJ and Respondent_RS were more pragmatic about how they share data.

> *"... I am fully integrated with Google's packages… and I love it. But I know the price there is privacy. I think when it comes to data and privacy in itself causes the most the most headache for me because I need to be careful what I commit to and not commit to in terms of data." (Respondent_JJ)*

> *"… privacy is becoming overrated… because of the fact that so much can be inferred through a "debated" guess. That whole concept about social engineering is exactly based on that premise. So for me it's like, when I look at the pros and the cons I rather [look at the] value that can be provided versus the restrictions and the lack of value. It's a clear choice for me… The value that I get from my device understanding my context outweighs the risks posed, and if I have an issue about it I can take steps to address it." (Respondent_RS)*

Respondent_JJ applied three factors to privacy: psychological privacy, online privacy and real world privacy. This linked the conversation to a previous concern, the infringement of civil liberties, highlighted by Respondent_AT.

> *"… there are 3 factors to it… There is the psychological privacy that we need. So there is that stuff you would share with people and not share with people. Then there is that online privacy and there is your*

> *real world privacy. So, for me for a long time I kept a very low key footprint in the online world. I came from a data security side when I was younger and learnt a lot there and I was like ok cool." (Respondent_JJ)*

Investigating these particular factors, respondents were asked how they perceived the statement: *"if you have nothing to hide, you should not care whether private companies or government agencies monitor and analyse online behaviour".* The respondent's reactions and feedback were interesting in that while they recognise the massive benefits that Big Data bring to organisations, but more importantly to their personal lives, they were very uncomfortable regarding governmental and organisational surveillance.

Respondent_TW marked the statement as hypocritical:

> *"You know that's fairly hypocritical in terms of saying "Everybody's life must be an open book" but then when people turn to your personal life you don't actually like it very much, do you?...this is deeply philosophical and prudential and perhaps a human right in a way, our ability to function effectively requires privacy… This is something fundamental to how we actually behave, it's built into our behaviour and that's not to say that things don't change and we can't benefit from that change, but at the same time you have got to respect the fundamentality of "hey if I want something", it's mine number 1 and number 2 it's part of what we have as a human right. Those are the two things that you need to have respect for Mr Organisation in the Marketing Space or the Google space or whatever it may be. You need to respect that." (Respondent_TW)*

Respondent_JJ felt strongly:

> *"I disagree with that, I think that people, just because you have nothing to hide doesn't mean you need to give out everything. I think that the intent for what government or big organisations collects, the intent behind the gain of that data is what causes the stress for me. So I am ok that they know who I am, what's my race, what's my preference in food and crap like that. There is a place in time for targeted media to come to my side and I appreciate that it reduces my Google searches. The government wants my information because they are trying to do something, track me for speeding, my tax money or something. I think*

> *it's the intent that makes me upset more than the actual information."* *(Respondent_JJ).*

Respondent_AT first introduced the consideration of ethical philosophies as well as the concept of education and the effect that it might have on the way that an individual would view his or her own privacy.

> *"This is where the ethics gets involved and if you ask "Billy Bob" coming from the wrong side of town and an educated person that comes from an MBA background you're going to get vastly different answers. So do I want to be monitored, do I have something to hide? If I have something to hide is it okay or is it not? What happens if I have nothing to hide now but I might have something to hide at some point* [in the future]? *So my ability to act outside of the framework that I have been born into which is societal control and laws and the Constitution..." (Respondent_AT)*

Both practicing attorneys – Respondent_DK and Respondent_NB – elaborated on Respondent_AT's concerns, stating that the concept of privacy is a very important human right and in terms of the South African Constitution, deemed as such.

> *"It's not about having nothing to hide. So remember privacy actually came into play because of constitutional rights and human rights - I have a right to privacy. That's my own personal right like I have a right to clean water. So privacy is my personal right, it's not to say from I want to hide anything from the government or other organisations but I should be able to decide what my data can be used for and how it can be used in order to exercise my right." (Respondent_DK)*

> *"We have the right to privacy enshrined in the Constitution but you can't really implement it, so what POPI does is that it puts all these obligations to protect consumer information to an extent, so you have to have safeguards, you have to be aware and transparent and know what you're doing, so it set this framework that we've never had before, so we've gone from zero to something." (Respondent_NB)*

Other respondents also mentioned the ethical philosophies of egoism and utilitarianism. Respondent_TW followed a philosophy of egoism:

> *"…my right is absolute as an individual, it is not weighed off against the rights of many, and many years ago we had a guy who taught me admin law saying: "You can take on City Hall!" You, an individual can fight City Hall and that's how the rights are enshrined. I think when you start becoming "The rights of the many outweighs the rights of the few" that sort of almost VETO to trample on the little guy. I can never support that, in my personal I have had a lot of that and I've never like it." (Respondent_TW).*

Whereas Respondent_NB questioned the philosophy of utilitarianism:

> *"The other extreme of the argument is, as private individuals what do we have to hide, maybe you cheating on your spouse, the government doesn't care about that, they are looking for terrorism and they only go into, filter through everything else and target the people that are actually showing certain tell tales signs in their behaviour and those are the ones they going to go for, but if you just doing slightly dodgy things but nothing that is of state or National Security then nobody is really going to pay any attention to you and therefore your privacy isn't really going to be affected and therefore why do you really care?" (Respondent_NB)*

Respondent_JJ paralleled the ethical considerations with The Stanford Prison Experiment that was a study of the psychological effects of becoming a prisoner or prison guard.

> *"…experiment itself taught us a lot - Stanley Milgram's experiment on prison, but he wasn't willing to take the responsibility for the repercussions of that experiment. So I think that in many ways Machine Learning is touching on those controversial experiments we used to do back in the day. Everybody wants to do it but nobody is willing to say OK cool but should we do it? And if we are going to do it am I willing to take the responsibility for it." (Respondent_JJ)*

It was further argued that while society yearns for ground-breaking innovations, human beings however rarely want to take the responsibility of controversial experimental technologies. Respondent_RS provided a real world example of the possible adverse effects of ground-breaking experimental technologies:

> *"…it will determine whether or not this will assist us, nuclear power is amazing, the abuse of nuclear power or even Chernobyl where you don't use it properly is terrible." (Respondent_RS)*

Respondent_JJ concluded with:

> *"… from a philosophical point of view, there is no right answer and there is only my point of view." (Respondent_JJ)*

Conclusion:
Respondents concurred that privacy is a human right and are apprehensive about the application of Big Data techniques to their everyday lives. However, the ethical theories of egoism and utilitarianism complicate the right to privacy.

5.6.4 Criticisms against Big Data

Respondents were presented with an incremental effect example: the Netflix Recommendation Experiment, already mentioned, where de-identified data was re-associated with identified individuals by cross referencing a de-identified database with publicly available resources accessible online. Once any piece of data has been linked to a person's real identity, the anonymity of the virtual identity is removed. All of the respondents were very familiar with this concept and identified several examples where they had experienced it in their personal lives.

Respondent_NB recognised the relationship between WhatsApp and Facebook:

> *"What I have noticed in my personal circumstances is that if you add somebody on WhatsApp you'll get the recommended suggestion of them on Facebook immediately even if you have no friends in common, so we're already seeing that type of immediate linkage of information happening in the background it's very hard to take yourself out of it, you do get people however who do shun social media to say; my privacy is more important that I will not be on any social media platform because I value my privacy." (Respondent_NB)*

Respondent_MS concurred:

> *"… something along the lines of Google and Facebook have more than 100 distinct characteristics which they can identify from a single*

> *browsing session which kind of helped them identify who you are." (Respondent_MS)*

Respondent_NB further elaborated, in that customers and consumers lacked the understanding that certain service benefits are derived from the incremental effect.

> *"So we want … all the benefits that we can get or that we can leverage, then we're very happy in that scenario but once we realise that we're part of that data set… giving us all of these services and efficiencies. People start thinking "Oh! But is that okay?", but I think that if there's a level of security,… a level of anonymization, those type of things bring comfort to people so they know although my information is in this data set they are protected…" (Respondent_NB)*

The issues surrounding the influence of Machine Learning, predictive analytics that influence online behavioural advertising and personalisation applications were put to each respondent as they raised concerns around discrimination, self-determination and the narrowing of choice. Respondents, although strongly in favour of artificial intelligence and Machine Learning, recognised the possible "blind-spots" within these technologies. Respondent_TW noted:

> *"…in essence… what is happening with automated decision making, is that it sounds good* [and] *we recognise that it requires a massive data* [set] *to be able to get it… but it's not really known where it's going and that means that there are blind-spots in the future… not-withstanding those blind spots, we take our hands off.. I am fascinated and I am excited by it because it starts moving into an AI world… into things that are absolutely I am passionate about." (Respondent_TW)*

Respondent_RS agreed that AI and Machine Learning is disturbing:

> *"In my opinion I would be worried about that. Here's why, I don't think that the average voter* [referring to Google's modified search results within the US elections] *does not have an opinion about who wants or who he or she wants to vote for, I also do not believe that something as subtle as who comes first in a search engine result can skew so dramatically that person's point of view. Are there risks? You know, if we take that to the extreme to say "you know actually I", let's say Mark Zuckerberg is a Democrat, which he is, he ensures that his results and search stuff on Facebook doesn't actually tell anything about*

> *Republicans. So is there a risk? Sure, but what that risk will be in terms of its magnitude and its impact, I don't think it's that significant, in my opinion." (Respondent_RS)*

Respondent_SW framed the concept of behaviour modification in terms of raising children with the desired social behaviours as well as the influence of reward based incentives:

> *"I always bring it back to the way you bring up children; if you raise a child you have to put rules and regulations in place to guide them in terms of decision making, now, if you use that as a concept or as an analogy then quite frankly, if a government wants to make sure its citizens are behaving in a particular way it then needs to reward for good behaviour and it needs to penalise for bad behaviour. But the only way that they can do that is if they understand your behaviour and we see it, without data or before this data explosion took place you saw it in the most rudimentary ways where you get tax rebates for certain behaviours… a typical one for example is: if you're married with two children well now that you've got a little family living together you get a family tax rebate but get divorced and you lose that tax rebate.*
>
> *…change of behaviour is very strongly linked to rewards and humans want rewards, it's like giving a child a sweet for being good or giving them a wooden spoon* [if they are not]... *The winners are the ones that use the data appropriately and get it right." (Respondent_SW)*

Respondent_HV acknowledges Respondent_SW's by reiterating the positive view on behaviour changes:

> *"So you're busy texting and driving and there's this thing on your screen, big flashing "STOP DOING IT!" …* [you will] *put the phone down. That could be a useful thing..." (Respondent_HV)*

However, Respondent_GS and Respondent_HV raised their dislike for customer specific marketing campaigns and framed these as desperate attempts by organisations to "squeeze the margin".

> *"Well you can argue that maybe they can change advertisements that are on sidewalks like bus shelters and as people walk past they'll change it to adapt the person they think you are as you read it you*

> *think "oh I was going to buy this coffee" it's like all desperate attempts to try and get the margin and squeeze every possible way that they can get your attention so that they can market the right thing to you, like market products or market people who you should vote for." (Respondent_GS)*

> *"I think that would just be one more situation where we do get influenced because you really think if you go to a shopping mall right now, it's years and years of retail experience and all of that being used… product placement… to even influence you right now…" (Respondent_HV)*

Respondent_NB speculated that as a society we will never be free and independent from these types of campaigns due to our inherent vulnerability to subconscious bias:

> *"So I think even if we take all of this out we will never be independent of it because we get that subconscious bias even from billboards without realising and even from the radio so this may just be subliminal messaging, we get that all the time it's in our face maybe this is just a more direct way because we interact with the computer so much and because technology has allowed it to be so tailored for us." (Respondent_NB)*

Respondent_AT and Respondent_DK equated behaviour modification to a level of propaganda as a control mechanism aimed at uninformed citizens.

> *"I don't think its behaviour modification… it is a propaganda race. At an ANC rally on the weekend, it's the same thing. People believe what they believe before they get told what to believe when it comes to these things, you either like Donald Trump or you don't. If you like him then you're going to like him, it doesn't what I try to convince you on you're still going to like him. If you like Hillary Clinton despite her indiscretions, if you like her and you've followed her for many years as a devout American patriot and you're a staunch American like most Americans are and you've followed Hillary Clinton for years, it doesn't matter. I try and target ads in benefit of my M.O. towards you really because you already like her. The problem is that the majority of people aren't thought capable enough to be aware of what things are, like what is*

> *the news, is it an information source or is it a control mechanism?" (Respondent_AT)*

> *"So where I am manipulating you to thinking something else, I mean, that is more of a propaganda issue than a privacy issue in my opinion unless you using my personal information to influence that views and your intent is different, then I've got a problem." (Respondent_DK)*

Conversely, Respondent_MS recognised that to deliver certain value added and "tailored" services, access to data as well as analytical inferences are required:

> *"…from a data analytics perspective this is particularly interesting… a really nice example of this is if you use Google and I use Google, you and I can sit side by side on two separate machines. You are open to your profile and I'll be logged into mine and I guarantee you if we search for the same thing we're almost guaranteed to get different results. So why is that? It's because your behaviour is tracked in terms of what you do, where you go, what you like, where you visit and it's the same for me. So they will tailor my results and your results to fit in line with what they think we want to see." (Respondent_MS)*

Predictive analytics were noted to have a useful application in law enforcement, national security, credit screening, insurance and employment. However, Respondent_TW was wary of this application:

> *"I am sitting on the fence at the moment. I think it's worrying, because I don't think we understand where we are going… Obviously it can be good for us, but… we don't have a perfect vision of the future around us… In a sense then taking your hands off the steering wheel, that's where I have a worry…*

> *… on the AI* [artificial intelligence] *side and the Big Data side of things we just need to be very, very careful and we need to be little bit wiser in how we approach these things. Sometimes going a little bit slow around things we don't understand. The future consequences of is probably what's called for. So hopefully the law will think ahead." (Respondent_TW)*

Organisations are reluctant to share the valuable insights gained by an individual's personal data and introduced the "access and exclusion" criticism. Respondent_MS raised the potential "litigation risk" that organisations would expose themselves to:

> *"I can't remember in the context of what this discussion was... somewhere along the lines of… there would be potential precursors to someone having a heart attack, let's just be theoretical, but let's say as an examples based on Big Data we've had indicators that show you had 5 events occur over a period of time, you're 80% more likely to suffer a heart attack… right now we're not sharing that with members and I guess we should be, whether that's ethical or not,* [and] *whether it kind of opens us up legally." (Respondent_MS)*

Furthermore, precursory indicators are at best an educated guess and subject to change as new data sets are introduced. Respondent_DK agrees that:

> *"… when we looking at predictive analytics, it's not always 100%,… you could never say in six months I'll change my diet and change things which can be a pro and a con. Maybe I can tell you about I don't know, I can change my diet and now I won't have a heart attack or change my job. I think whether they should or should not tell you is an ethical issue, not a privacy issue so if they going to sit on information like that, they got the ability to be able to disclose it or not disclose it. I know what they do up till now is they'll advertise you know, if you've gone for your annual medical check-up. You know like those at the banks that can predict.... they use it for marketing purposes and the intention of the company is to get the best out of you as opposed to looking after your best interests you know." (Respondent_DK)*

However, access and exclusion touches on various ethical philosophies and is yet to have a defining solution:

> *" …let's say as an example we've picked up something that says we have some key indicators and we believe you're going to have some type of threatening cancer that you'll develop in the next three months. Cool but what happens if it doesn't happen? Now it's a case of the mental distress we've put you through and all the legal bills you had to put together to put your will together. So again… I certainly do believe that the value is there but I think there are still some interesting*

> *problems around it that probably still need to be solved." (Respondent_MS)*

To summarise the respondents' views, Respondent_RS proposes that the level of risk a customer is willing to take on is directly proportional to the value that will be derived from the service:

> *"…* [the] *value will directly be accountable for the amount of risks I'm willing to absorb and that impacts directly the amount of data that I'd be willing to share. This is why for me it's all about value, the conversation needs to be about value. If you're asking me to share my data, the story that you need to tell me about the art of what is possible must entice me then to say, I want to free up my time to spend with my kids because then I don't have to worry about having to go shopping and I've got that taken care of. My fridge understands that based on ID tagging of all my food in it that - oh I'm running low on milk so it sends out order, pays for it and there's a delivery service coming. Poof! There's my food. Can we do that today? It's just about being able to build those ecosystems and having them interact." (Respondent_RS)*

Conclusion:

Respondents were of the opinion that most of the criticism against Big Data should, in certain aspects, be regarded as a "necessary evil" as the rewards outweigh the perceived risks. Additionally, the level of risk a customer is willing to take on is directly proportional to the value that will be derived from the service.

5.6.5 The efficacy of privacy legislation

Current legislation models are created with the ultimate goal of permitting only legitimate procession of personal data. However, as with Big Data, the efficacy of the legislation is questioned and several shortcomings acknowledged.

Respondents overwhelmingly conceded that privacy laws follow practice.

Respondent_TW notes that this is a historic problem:

> *"The law would tail far behind the practise; I don't think that has changed when it comes to privacy… This is about technology, before*

> *you know it, it's so deeply embedded people haven't even worried about what the legal practise is* [and] *now we're catching up, but* [the law] *catching up very quickly. Because it's a deep concern there is the academics running ahead and anticipating where the technology is going." (Respondent_TW)*

Respondent_NB agreed that technology innovation outpaced security and legal mitigation:

> *"From my experience often tech innovation moves very fast and security and legal risk is sometimes an afterthought." (Respondent_NB)*

Respondent_RS concurred that legal practice needs to adjust itself to technology advances:

> *"This is why you'll find, and it's going to be a bigger problem moving forward, technology will always outstrip, in terms pace, legislative frameworks. Case in point, take Uber, nobody knew how to deal with Uber.*
>
> *So you'll find that the law will need to adjust itself based on the advances in society that has resulted through advances in technology. POPI, POPI got brought about because of the advances in tech and web and information and data "Oh we need to do something about privacy and peoples information you know so let's put POPI together" right there." (Respondent_RS)*

Respondent_HV related this statement to the rate of change in technology innovation, whereas Respondent_DK stated that the law will inevitably always need to catch up:

> *"And I would agree with that because I don't think…if I just look at the rate and pace of change in the technology we work with; to keep on putting legislation in place, like to handle future cases and all that, it's near impossible hey." (Respondent_HV)*

> *"Laws in general are not keeping up with it, they can't. The amount of time it takes to create a law. It's taken POPI good 11 years to get enacted. Jeez, what has happened in 11 years. So that's why what they try to do is create principle legislation as opposed to hard and fast rules. So POPI is based on principles. There's a principle of further*

> *processing and processing limitation and then you can adopt it into your company as you will." (Respondent_DK)*

Considering "Informed Consent" respondents agreed that individuals neither read nor understand privacy policies as the language is deemed ambiguous and time consuming. Respondent_SW stated that:

> *"Nobody really reads them, I mean let's be honest and that's a classic example of companies having to spend money to demonstrate that they are compliant, and all they do is that they shift the accountability to the consumer and the consumer doesn't read all that rubbish so for me there is the element of absolute waste, but that's my personal opinion. The bottom line is if you're going to buy a product from a company or you're going to use an application, there are Terms and Conditions that go with it. It goes beyond me what Terms and Conditions." (Respondent_SW)*

Respondent_JJ suggested that he doesn't even read his employee contract and that businesses spend too much time looking for loopholes:

> *"My buddy is a corporate lawyer and so he asked me how are you guys with POPI? I was like we are doing all of this and this and this. He tried to understand how it affects some of the data we work with, and I said to him 'look it's simple'… But if we really want to push the edge, we need to understand that sometimes the rules itself that is there to protect us, has so many loophole in it that we are spending a lot of time working around the loopholes instead of pursuing what is really great. I think that everybody is as aware as we are with what is going on in data and what companies collect. My mom has no clue how much data Facebook has about her.*
>
> *I think the concern is for the people who don't know what does that entail. My mom and your mom; that is my issue." (Respondent_JJ)*

Respondent_AT agreed that consumers blindly accept terms:

> *"…read that you'll understand exactly what your system is going to be offering up to the manufacturers of the software in terms of information to allow the software to meet its end. Now we blindly just accept these whatever "click, click next" and we use Apple and iTunes and Apple*

> *music and Google services and whatnot; we use our phones and we don't really know, I mean there's a whole bunch of the average on the street reads or not for that matter, we just assume that the company is operating honestly." (Respondent_AT)*

Furthermore, the principles of informed consent were questioned as well as the level of education required to understand what one is consenting to. Respondent_NB mentioned that:

> *"…consent is a misnomer but I think you really should have transparency and awareness so the consumer can then be aware of what's happening because to say that you're getting consent by ticking some box is not real consent, I don't think that meets the standard that POPI and other legislation have set out." (Respondent_NB).*

Respondent_JJ questioned the culture change that is required:

> *"Yeah you could say there is culture change, but I don't want to see it as a culture change because I see needs to be a part of the basic tools you get as a child when you grow up and we're not getting that. I think it's an educational problem. This is beyond culture. Culture is not going to fix this but if we can change culture everybody can learn to say NO. If it says YES/NO and it's more than one sentence say NO, that I can change we can become a culture of YES clickers NO clickers. But I think it's an education issue, it is definitely education. I've seen stuff where, and I consider myself technologically savvy where I even went like: "Maybe I"m not'." (Respondent_JJ)*

Respondent_AT and Respondent_NB examined whether there were available alternatives to lengthy ambiguous terms and advocated a picture based format that would communicate concepts across cultures and languages.

> *"Legislation could be put in place which could minimise the complexity of Terms and Conditions because an End User License Agreement does not need to be 70 pages... I don't know if you've ever seen the iTunes one? That is something that no man or beast should ever have to read… I don't think the companies are only* [ones] *at fault… it's a 50/50 thing… consumers are as much to blame as the organisations... but they have an end… it's as much their fault for exploiting human*

> *weakness but it's as much our fault for bypassing our own individual controls." (Respondent_AT)*

> *"So they* [legislators] *are trying to fix the problem on a global level but it's a global issue being that the laws don't always protect privacy in the way that it should… the EU-regulation is trying to do is making it picture focused, so they will be like a eight symbols, either green or red, which tells you quickly what this privacy policy is saying without actually reading the policy, so you'll say I accept when you already know are they going to be sharing with third parties, is it going to be same cross border just by looking at those eight pictures which represent various privacy principles." (Respondent_NB)*

The ambiguity of PII remains wide-ranging; therefore, organisations use various methods of de-identification (anonymization, pseudonymisation, encryption, key-coding, data sharing) to distance data from personal identities. Respondent_MS again referred to the lack of education and understanding attributed to privacy legislation:

> *"The one thing that I think, you say that PII is not well identified and in terms of if some of the acts that's probably true, but the one thing that I think that the regulations helped a little bit with is that consumers don't really have a good understanding of privacy in general, is my gut feel." (Respondent_MS)*

Limiting privacy exposure though data minimisation was deemed as restricting the value and premise of Big Data. Respondent_GS drew a parallel between data minimisation and the practicalities of cancer surgery.

> *"…when you want to operate on a patient and you only cut out 2% of the cancer and you can't remove the whole thing it's like "Come on?" If it's the company's data no one's got anything to say about it because the company is saying, "here's my data", I'm saying "well if you only showing me 2%. It's like - look through the keyhole and try and solve the crime - but I need to go into the room where the scene of the crime was with my investigative cap on and I need to see what is going on but if you close the door and say 'look through the keyhole and solve the crime' - my chance of solving the crime are going to be a lot less… I don't buy any of that." Respondent_GS*

This statement leads to the intent or purpose of initial data collection and whether subsets of data could be closely linked to the original scope of engagement. Respondent_DK builds on the original intent:

> *"…when I asked you for your permission. Could my subsets of a and b actually be related? So the law basically states if it's related then its ok, I'm not completely moving away from the scope of what I actually wanted to do with your data. So let's say as an example, I originally gave you my data because I wanted your loyalty card. As part of the loyalty code, you going to tell me across the different stores which products I'll potentially like to buy. But if you going to use my data for marketing a funeral policy, we going to have a problem cause that's not why I originally gave it to you so it depends on how far out that it is so what the law basically says if its reasonably the same as my original consents and the purpose that you originally wanted to use it for then its ok but I need to be able to showcase that its reasonably the same." (Respondent_DK)*

The conclusion reached by Respondent_DK was closely aligned to the principles outlined by the South African POPI Act.

The intent of the Big Data organisation or service provider is a recurring theme throughout the interviews conducted.

The relentless improvement and expansion of technology capabilities and information globalisation was noted as a contributing factor to the ineffectual application of privacy legislation. Respondent_AT agreed with this factor:

> *"No,* [legislation's response to globalisation] *is way too slow,* [and] *especially in developing countries legislation is too slow." (Respondent_AT)*

Respondent_NB cited an EU case:

> *"…in Spain where Google was ordered to remove certain search results because they said it was non-compliant with Data-Privacy law because it was excessive, irrelevant, outdated etcetera ticking all those boxes, and Google complied with that judgement in the EU but it has no practical effect because if you're sitting in the EU but you use*

> *Google.com or Google.co.za, you're going to get your full search results including the omitted search results." (Respondent_NB)*

Respondent_NB further explained that:

> *"From a legal point of view, the law has tried to implement and make a court order of what the law says "Google you have to comply with this because the law says xyz", practically the way technology works and the way we are as a global technology community, you can't really always implement those in a practical manner so it actually has no practical effect having a court case in the highest court in Europe." (Respondent_NB)*

Respondent_NB offered alternative solutions to the latency of legislation. Specifically, ensuring that products and services involve legislation at every step of the research and innovation (R&D) phase. This would mitigate significant risks and allow services to be more secure from the start:

> *"So you'll get asked to rubber stamp a product that has already been developed without making sure that its legally compliant or that something might happen, and you'll say if you actually spoke to us when you were in the R&D phase and hadn't built the product yet we could have identified some of these risks, we could have built that into the product because it's very hard to reverse engineer. And then you come into the situation where you have to work with what you have and make it work, and I think if a few steps back when we start having those questions about risks and how to mitigate against those risks, then those would actually be built into whatever product you develop and then it will become more secure rather than adding a patch at the end to try and make it secure." (Respondent_NB)*

However, other respondents disagreed with this premise as they were concerned that involving legislators in the research and innovation (R&D) cycle would stifle innovation. Respondent_RS marked privacy legislation as:

> *"…restrictive and it actually takes away from the value. I understand that there needs to be some guidance and frameworks in place for addressing things to understand the need for it, but I think that it needs to be flexible and robust enough and handle things on a case by case*

> *basis to ensure maximum value add as opposed to restricting…" (Respondent_RS)*

Respondent_TW questioned the qualification of innovation:

> *"…and I think if it's got a component of longevity, sustainability… and ethics around it then, you know, no you are going to have to strike that happy balance and I think that's probably the better way to go." (Respondent_TW)*

Respondent_HV stated that companies would be reluctant to share intellectual property discussions as these opened them up to significant business risk.

> *"So I don't know if it would make a difference if the legislator or whoever sits in the room with you know, the guys coming up with breakthrough technology. First of all I don't think the tech companies would do that because the moment you do that, you know, it's… your competitive advantage is gone right then because you've shared what you're going to do with these guys." (Respondent_HV)*

Conclusion:

Although the legal premise of Privacy Legislation is sound, it has failed to keep pace with globalisation. With the relentless improvement and expansion of technological capabilities as well as the changing ways in which individuals create, share and use personal data, privacy and in some ways security legislation frameworks have been unsuccessful in meeting their goal.

5.6.6 Personal Data Store

The premise of moving data collection from the individual as a participant only, to the individual as the centre of personal data collection, management and use was generally welcomed.

Respondent_TW agreed that it made sense:

> *"…in theory. From that perspective I am completely supportive of it… it's an interesting thing because it almost gives some tangible aspect to your data/to your personal information." (Respondent_TW)*

Respondent_JJ liked the premise:

> *"…I think the idea is a good idea, but the actor is in control so the person is now in control." (Respondent_JJ)*

Respondent_MS agreed that it would put a lot of power into the hands of the user:

> *"…it doesn't fully address the problem because it helps a lot right, don't get me wrong, I actually really like the idea, I think for me the biggest gap still is the point of once the data is in possession of the company as an example, what stops them from transferring it somewhere keeping a local copy of it and selling it away." (Respondent_MS)*

Respondent_RS believed from a theoretical standpoint it would:

> *"…even things out. It should be opt-in versus getting it by default and opting out, so yeah I agree with that because you sign up for a service, when you take something out you sign up for it and when you sign up for it you opt-in and you understand what comes with that and you decide when you want to opt-out or not." (Respondent_RS)*

The concept of a PDS introduced the intellectual coherence and property base theory of personal data. This theory resonated with Respondent_RS:

> *"Yes, I do think that would be cool, because then I can find creative ways of potentially monetizing that. So yeah the ability to choose, again, is paramount for me, yes its very valuable for me." (Respondent_RS)*

Respondent_TW provided a comparison:

> *"So almost think of it like - it is my asset and I will decide who I will give it to and when I'm going to take it away and if I lend it to you it's given to you in a certain condition etc. so it makes sense that it's almost becoming a piece of either tangible asset or intellectual property." (Respondent_TW)*

Respondent_SW saw the possibility of commercialisation:

> *"So instead of companies making money of my data, give the consumer the right to make their data available to be sold. And then ask what data I have about myself that I want to add into their pot and*

> *pay me for that. Sure. I'll even tell them my behaviours too." (Respondent_SW)*

Respondent_JJ and Respondent_AT made a comparison between personal data monetisation and music royalties' business models. This is an interesting comparison as it also refers to Data Rights Management (DRM) of meta-data tagging.

> *"I think ethically it's the right thing to do. If you are going to use my data and you are going to get something valuable out of my data, there should be [compensation] like royalties. I think that's the right thing to do, we do it in music, we do it in art, we do it in book, we do it well everywhere. Why can't we then do it with data?" (Respondent_JJ)*

> *"The minute you monetise my information then there are two things that need to happen; either you need to stop what you're doing or you need to give a slice of the pie… give me a slice of what I am then I'll let you do whatever you want but if you're not giving me a piece of that monetization…" (Respondent_AT)*

Although the premise of a PDS was welcomed, its technical feasibility was questioned. Respondent_TW showed apprehension with regard to one system's ability to capture all-encompassing data.

> *"…the difficulty that I have with it of course is this, which is that our personal information is so broad. The ability to capture it all in one place and sort of go hang on I've got a data store of my personal information, now obviously you can and just think of it, if I go to gym this morning at the gym I go on the treadmill and then I stop I have a smoothie afterwards and on the way back home I stop at the coffee shop. So each one of those location points plus what I have collected is actually potentially personal information." (Respondent_TW)*

Conclusion:

The premise of consumer empowerment with the individual as the centre of personal data collection, management and use was universally welcomed and introduced the intellectual coherence and property base theory of personal data. However, its technical feasibility was questioned as regards the system's ability to capture all-encompassing data – both dynamic and static.

The ability of customers to share their data selectively without disclosing more data than they wish to – selective disclosure – was compared to an "opt-in" model. However, Respondent_TW revisited the data minimisation criticism of privacy legislation as well as the difference between static (once-off) data and dynamic (continuous or live) data:

> *"Unless you are going to have a data store that is literally updating all the time, it's feeding into your phone and your phone into your data store and then I store it at a server at home or something like that. Then I say I am now going to release this to Microsoft. The reality is you are going to potentially not be able to benefit because you are, that real-time personal information is then not getting into the system. Which means that the services that benefit from real-time or close to real-time data will not be able to accept that." (Respondent_TW)*

However, once the researcher introduced the view that the "selective" data that is provided to an organisation would be in data terms be "perfect" (referring to data veracity), the respondents were more welcoming of the selective disclosure. Respondent_RS stated that:

> *"… it will minimize the technology layer and the Machine Learning and all of the assumptions. Now you apply current technology to accurate Meta-data, you're going to get a much better result." (Respondent_RS)*

Supported by Respondent_SW, Respondent_JJ argued that data cleansing is a real operational problem for many organisations and a very costly one at that:

> *"I would argue that 80% of this whole data thing is just like shift work or like they call it janitorial work of just cleaning, preparing, worrying about missing values, incorrect data points and all sorts of inconsistency in data and that's what 80% of your time in many of these big projects and before you even get to the luxury of saying 'then let me apply this tool, let me do some Machine Learning." (Respondent_JJ)*

> *"I absolutely think that the biggest challenge companies have is good quality data, so if you think about it, if you take a consumer and you gave them the ability to transact with you directly no human hands have touched it other than the consumer." (Respondent_SW)*

Building on this Respondent_JJ, again supported by Respondent_SW, noted that it would not only reduce operational costs but also increase strategic efficiency:

> *"So yes, it doesn't just reduce your operational costs it also increases your strategic efficiency. We are using data to make decisions, you don't use anything else to make decisions data should be the fundamental thing. So you increase the quality of that, you don't just decrease your operational costs but you also increase your probability of success going forward. Even if you do bad science with that data, you are still going to get a better answer. You are definitely going to get better answers." (Respondent_JJ)*

> *"I think that companies that get this right are going to differentiate themselves in the market. Not their products. Not their services. It is what they do with their customer data and their knowledge of their customer. How they improve their business processes, with the knowledge that they've got. How they bring the cost of their products down, how they can implement loyalty programs, rewarding the right behaviour or financial management behaviour – because that is also possible." (Respondent_SW)*

Conclusion:

The concept of selective disclosure supports the ability of customers to share their data selectively, without disclosing more personal data than they wish to. However, data veracity is required to mitigate the data minimisation risk within the model.

Respondent_TW and Respondent_RS agreed with the proposition of control over purpose and duration of data and further proposed that should this proposition not be included; personal data would lose its intellectual property coherence.

> *"For me, and that's where I want* [a PDS]… *if I think you've got my information you must be able to identify it or not with certainty and if I ask you to delete it or change it you can do that. That's how simple this thing should be. Then I'd feel safe and secure. Until that has happened we are all going to be sitting at this sort of quandary of the unknowing and it's not a great place to be for me as a consumer and for me as an individual to not know." (Respondent_TW)*

> “[For example] *I say ok Medi-Clinic you can take my genetic information and let’s pretend it is just going to be used in some kind of Big Data thing. Great, once it’s done it’s done. I can now say give me my information back. The benefit and consequence of me having given it is out in the world, I can’t take that back. So I lose ownership and proprietorship around that in a sense.” (Respondent_RS)*

Conclusion:

Personal data would lose its intellectual property coherence should the premise of control over purpose and duration not be included within the PDS model.

The respondents struggled with the concept of signalling where individuals express demand for goods or services in open markets not tied to any single organisation. Some grappled to understand how signalling is different from traditional tailored marketing. However, when the researcher explained the signalling in terms of a “pull” action rather than “pushing” marketed products on to a consumer, the respondents received the concept as positive. However, on the question of whether the “pull” action would result in higher revenues for an organisation, the respondents were split.

Respondent_AT equated the concept to a “qualified” lead:

> *“…for them versus a sales person having to qualify the lead on the company’s behalf. If I’m selling Microsoft solutions, it’s much easier selling that product to a customer who has already decided what they want. All the need me for as a sales person is to confirm their already made decision, so in this example if I signal as a Data store that I’m in the market for “this”, I’ve already qualified the lead, I’ve already decided and I don’t need to be convinced I just need someone to call me, then means their sales efforts are more accurate or more efficient.” (Respondent_AT)*

Respondent_JJ mulled over the point that customisation of product and services was expensive:

> *“The interesting thing about that, like in my mind from a company perspective in terms of sales and revenues I would question whether it would drive it up or down. I think it’s probably the potential for both in a way I think the contour or the downside in terms of where you may*

> *lose money is that you have to know start individually tailoring something on an individual level." (Respondent_JJ)*

Conclusion:

The premise of signalling – "pull" action rather than "push" – was not deemed as a foundation block of the model, but questioned the premise whether the effect of customisation would increase costs rather than the desired cost reduction.

The terms "security" and "identity management" were used interchangeably by respondents and were not viewed as mutually exclusive. As part of the same process of interacting and transacting, organisations in both the private and public sector need to be confident that the person they are dealing with is who they say they are as well as that the security protocols are of the highest standard. Respondent_MS agreed that identity management and security should be entrenched at a foundational layer:

> *"From a consumer perspective… I would like to know that my signature belongs solely to me so in terms of like non repudiation, there's never any discussion about whether it's me or whether it might be me… the general management around that identity needs to be quite tight… [a] rigorous process on how those identities get issued how they get revoked… perhaps upon death or whatever so that you never really end up in a position where identity fraud is a problem… The technical complexity around that is, would be quite something. Interestingly enough I mean like whatever it would be it would have to be very math heavy it would be very crypto heavy, so in the question there whether or not there is anything sufficiently good enough right now to do that?" (Respondent_MS)*

Respondent_DK concurred and argued that it:

> *"…depends on the actual authentication, if it's single or multi factor authentication on security perspective as well because there is certain security controls that you can put into place that can make sure that [a person] is who she says she is to an extent. So things like biometrics with a card, with a password, whatever the case is, you can kind of guarantee that I am who I say I am, where I only use a username or password, you can't guarantee it so I think that's going to be a component of It. In terms of whether it would work, it would work if I*

> *put the right controls in place and I've really honoured it like the bible to make sure that everything has been put into place. In terms of keeping up-to-date with accurate information, I can only confirm it's accurate if I'm able to update where. Have I moved, have I changed my number, have I changed my job. So it needs to be able to be updated with me otherwise the information you holding is not up-to-date and accurate cause you got it six months ago. So I think it depends, in my mind." (Respondent_DK)*

Respondent_SW however went a step further, believing that organisations are not doing enough to ensure identity management:

> *"It is quite the opposite. In the world of identity and access management if that is done correctly, you can create safely and security for your customer – not irritate them. In fact, I have yet to come across somebody who doesn't appreciate it when you ask them to validate who you are. You know, you phone the call centre, actually I can phone the call centre now and give them my brokers ID, not ID number, their intermediary code and they don't check to see if I am really that broker, and I can ask to see my broker's portfolio clients." (Respondent_SW)*

Respondent_GS introduced the application of Blockchain Technology as the preferred form of authentication and security layer as there is instant non-repudiation and that transaction cannot be reversed or changed.

> *"The Blockchain universe is an interesting concept because it is transparent but it's still hidden and only you if you've got the keys can open the data, the data is there and everyone can literally go look on the Blockchain and see if it's on there but its encrypted but everyone can see it and you're never going to ever unencrypt it." (Respondent_GS)*

Respondent_AT agreed that:

> *"…MFA–Multi-Factor Authentication is a good standard for companies that have more sensitive information that people need to access, I prefer the concept of Blockchain because personally when it comes to Identity Management if I can be judged in a group of my peers as to whether I am who I say I am, so there's non-repudiation there instantly*

> *and the fact that that transaction cannot be reversed or changed, that's much better." (Respondent_AT)*

Respondent_JJ approved the technical viability of Blockchain technology but compares it to the failed adoption of the peer-to-peer (P2P) platform:

> *"…P2P is a beautiful example of Blockchain right? The truth is that P2P as a tech has failed completely, it functionally complete and it works as it's supposed to but the general adoption of it is so poor that it never got off the ground so from that perspective it concerns me that even though it's a viable solution and technically it could work, adoption fails." (Respondent_JJ)*

However, Respondent_GS disagreed and argued that the adoption of mobile phones, Machine Learning and artificial technology will ensure critical mass within adoption.

> *"These broad based technologies like Machine Learning and Blockchain which is like, it's not Bitcoin, Blockchain is different it's the other line of technology which does use some Machine Learning as well in its implementation which uses basically distributive file systems and cloud infrastructure in an amazing way. These things are significant technologies, Blockchain is going to be significant, Machine Learning is going to be significant and Big Data is a consequence of stuff and so a portion of Big Data will be useable and a portion of it will be significant but it's not going to be in the same league as the other two things in terms of big stuff that's happening as we look forward in time, I mean Big Data will be around for just as much as data has been around forever and it's just bigger because there are these mobile devices." (Respondent_GS)*

Respondent_NB recognised that Blockchain technology is cutting out that trusted intermediary, possibly eliminating a possible institutional void.

> *"…there's also the privacy debate on Blockchain because what Blockchain is doing is cutting out that trusted intermediary, because we're saying we all can see what is going on and we don't need an institution to tell us what's happening." (Respondent_NB)*

This was echoed by Respondent_RS and Respondent_GS:

> *"So the bold view of Blockchain as a premise is that it eliminates the middle man." (Respondent_RS)*

> *"What I think it does is take out the middle man, if I have a song I don't have to use a music distributer, so that's part of the reason why Blockchain technology could work in different areas is because you don't need to have distribution channels. It's a much fairer, the artist actually gets the value, and they have control of the keys on the Blockchain. Their work is authenticated and verified on a Blockchain so you know it, it doesn't matter what happens you are the owner of the technology because you've got the... Blockchain has recorded it. so there's no fudging of the boundaries and I think it makes the payments and the right for the royalties to those creative people it makes it more direct so that's what makes it happen" (Respondent_GS)*

In summation, Respondent_GS maintains that the premise of Blockchain technology is not so future-based as some critics argue.

> *"…these smart kids from South Korea and Russia are collaborating and will use alternative protocols that will allow people to keep their own data on Blockchain type structures and that's going to challenge the current status. I don't think it's fair that people know all these things about you but there's nothing you can do about it for the moment so you just deal with it and when you get to the point where you've got your own data and it's yours and there are rules that protect it for you then there is a bit of space for you to operate. Banks can have credit card transactions on Blockchain and you will be able to have access to that, the Bank might have limited access to that so then it's your data and you can sell it to who you want to." (Respondent_GS)*

Conclusion:

Identity management and security should be entrenched at a foundational layer of the PDS model. The introduction of Blockchain technology is a viable solution for various technical shortcomings of the model.

Respondent_GS argued that the ability to move personal data from one provider to another using standard data formats and interface protocols results in a moot point as the transfer of personal data is no longer required via the Blockchain platform.

> *"Well if it's on the Blockchain it will never go away it's there permanently forever, for the rest of time it will never be removed you can only insert data on Blockchain you can't delete it." (Respondent_GS)*

Conclusion:

The ability to move personal data from one provider to another using standard data formats and interface protocols results in a moot point as the transfer of personal data is no longer required via the Blockchain platform.

Respondents viewed the accountability and enforcement of privacy as well as a PDS in two ways. Firstly, some respondents felt that the accountability for protecting and securing personal data sits with the 'receiving' party – in this case the organisation. Respondent_TW stated that it should be enforced under self-regulatory guidelines and legal mandates, both backed by comprehensive auditing:

> *"Similarly, with POPI, they have made it the person who is receiving the responsible party, which generally is the company, they have made them responsible. It is probably the model that will continue, because that's where the force of money sits. That's where the power sits, that's where the administrative organisational skills sit. That's also the easiest way to effectively govern, so it'll probably continue in that space. The reality though is that unless we as individuals buck up and start understanding what's going on. The fact that it's sitting with the company has a responsibility or legal responsibility becomes an irrelevance. I can't test it." (Respondent_TW)*

Respondent_NB and Respondent_SW agreed:

> *"…to an extent from the law and from the government but I also think there's a big space for corporates because as we've seen the government can sometimes try their best with the law but they get outdated very quickly because of the processes it takes to actually bring something to the foreground or they practically can't implement*

> *certain things no matter what the intent may be behind that law, so although they have a role to play in protecting their citizens, they can't do it alone and I think there is a space for good corporate citizens to say when we have your information we are not abusing it and if you have any issues with that we will either stop using your information or change the way we are using your information depending on the circumstances. But there is a certain level where big companies come into it, especially big global tech companies because they often drive these things." (Respondent_NB)*

> *"…again here, I think the regulator can play a role. Ok, if a company is accountable to ensure that they identify and that they restrict access to the individual's information. It is their accountability as far as I am concerned. And it is the education of the consumer that if a company does not validate and you're not feeling safe that company is validating that you are who you say you are, don't deal with the company." (Respondent_SW)*

However, the efficacy of institutions to perform this accountability function is contested by Respondent_MS:

> *"So ideally it should be the regulators, especially if you look at is from a POPI perspective right? So the regulators would be responsible for your auditing and basically ensuring that the company is correct and if they don't then they should be fired and instituted against. I mean it's kind of and interesting one mean if you think in the same way if you look at data breaches globally right? Have you ever seen anyone prosecuted or convicted? Why is that?" (Respondent_MS)*

Respondent_DK agrees with Respondent_MS that legislation is only as strong as its implementation; hence effective regulators are required.

> *"It's all going to be as effective as the implementation so I'll give you an example. when the consumer protection act came into law we all thought it's going to be this big bang consumer issue but the regulator is too snowed under and what they assigned for the regulator's establishment was too little so they really on the back foot whereas you take the Competition Commission and just the other day the*

> *Competition Commission handed over another big fine. That's a strong regulator." (Respondent_DK)*

Conversely, some respondents feel that government regulators and legislation should not take ownership of this PDS model and that Blockchain technology provides for the ultimate accountability.

> *"Blockchain offers the ultimate in accountability and verification from a compliance point of view so as we speak the top 6 Banks in South Africa are meeting on a weekly basis to discuss how they can use Blockchain technology because they can't afford to ignore it so a lot of the transactional processing on the exchanges like the stock exchange will move to Blockchain and this will take out a lot of the middlemen, there will be no need for them." (Respondent_GS)*

Conclusion:

Reliance on regulatory institutions merely extends as far as the effectiveness of the institution does. Conversely, government regulators and legislation should not take ownership of this PDS model. Blockchain technology provides for ultimate accountability.

Two additional concepts were jointly raised by respondents.

The model assumes that all individuals are 'knowable' in the realm of personal data and privacy, whereas the opposite is true. Respondent_TW argued that organisations lack the required skills to deal with personal data and privacy.

> *"That's what this is about, and he's going to protect it while he's got it. It's not difficult, but the problem is there is two things that happen in that space, there is a technical aspect (the people who have got to get it right in order to do that) and there is an operational aspect to it (which is the security, the personnel, the training and all the 'if somebody breaches this, what is the disciplinary approach?' etc.). There is a lot of work that people have got to do, there is muscle usually within business organisations particularly in South Africa that is massively underdeveloped that needs to get developed." (Respondent_TW)*

Respondent_MS agreed in so far as visibility and awareness were considered:

> *"…around privacy issues in general, I think that people don't fully understand what their data is being used for, and your quite technical I mean you have quite a big understanding on Big Data obviously and kind of know what is being done but if you had to ask a general* [Organisation] *member as an example, right just say to them "right your data is stored at* [Organisation]*, do you have any idea of what data they hold on you and what they could process?" you'd probably just get a blank stare right? But the truth is that when you think about it especially in our context, if you think about, let's assume you're across the board in terms of customs you've got your live data, your health data, your credit card your GPS data potentially so we kind of know where you live where you sleep, we know that value of your assets we know what you spend money on. In terms of the amount of personal information we can glean on you it's astounding and people either are oblivious to that fact or they are just not aware or they don't care, I'm not sure which it is. For me it's a subject that needs a lot more discussing and needs a lot more visibility and awareness." (Respondent_MS)*

Respondent_JJ, echoed by Respondent_TW, recognised the generational and perhaps cultural gap in the understanding of Big Data privacy.

> *"I think the concern is for the people who don't know what* [Big Data and privacy] *entails. My mom and your mom; that is my issue.*
>
> *We don't teach that at school, we don't teach data at school, she* [a team member] *has an email address with password, the email address is her full name and surname with her date of birth at the end because her name and surname was taken so she actually just typed 19860what what her birthday. Ok now there immediately just an ignorance, this is not a cultural issue this is an ignorance issue nobody told her that." (Respondent_JJ)*

> *"It's huge, the world has shifted too quickly and the individuals are being left behind. That's the difficulty and there is going to be a generation which is going to be left behind and hope that the next generation will be born into it. So it could be a generational thing at the end of the day." (Respondent_TW)*

Respondent_JJ again proposes that society is in dire need of a cultural change that includes data privacy as a basic tool within early childhood education.

> *"Yeah you could say there is culture change, but I don't want to see it as a culture change because I see needs to be a part of the basic tools you get as a child when you grow up and we're not getting that. I think it's an educational problem. This is beyond culture; culture is not going to fix this but if we can change culture everybody can learn to say NO. If it says YES/NO and it's more than one sentence say NO, that I can change we can become a culture of YES clickers NO clickers. But I think it's an education issue, it is definitely education. I've seen stuff where, and I consider myself technologically savvy where I even went like: "Maybe I'm not"." (Respondent_JJ)*

Respondent_GS found that in his experience young adults in the United States were very informed and took their personal privacy very seriously.

> *"Yes I went back and that's where I did the Master's degree in Computer Science so my knowledge of it is current and recent so I didn't do it like 25 years ago and the reason why I mentioned is because the students I was studying with, and I was still working and I had my job but I was doing it remotely so I was like studying full time, I was working full time, I had kids and what I learnt in the process is that a lot of young kids like kids in their twenties had no Facebook accounts and they said that they don't want one but the lecturer had said that this is how this class communicate, you've got to setup a Twitter account and you've got to do this and that. And the kids go no we're not going to do that we have the right to choose what happens with my data and they were very informed." (Respondent_GS)*

Conclusion:

The PDS, and organisations in general, assume that users and consumers understand data and privacy.

Lastly, considering the multifaceted dimension of personal privacy, all respondents mentioned government and organisational "intent" as the deciding factor when reflecting on their personal privacy. Respondent_TW questions:

> *"…whether or not people are with that, you know when it's serving a specific purpose as opposed to a nefarious purpose. I mean I guess I can take a little side break here, this stuff is very, very interesting at the end of the day because I think you can almost link it back to Oppenheimer and sort of to the atomic age. This stuff is great for us, it's how we end up using it." (Respondent_TW)*

Respondent_MS and Respondent_JJ agreed it was not the fact that an organisation had the information but rather their intended usage.

> *"That's interesting, but I think it's also interesting because the nefarious element of it, the bad element is that once you're no longer anonymous online, that sort of darker side of the internet become a bigger problem. So I want to know more about you I know just have to target a few key services that may potentially have access to your data and the truth is if you think about, is that almost everyone today will use Twitter or Facebook or Google, which really means that I have to compromise one of three services to you to understand everything about you and that a privacy concern and I guess from a Nation state perspective it also means that it's a lot easier for Nation State to get that information." (Respondent_MS)*

> *"Because that's the problem, it's the intent, it's not the fact that they have the data." (Respondent_JJ)*

Respondent_JJ further elaborates that transparency on the intended use will likely shift his view on sharing data.

> *"Do I agree that the information should be available to the government? No I don't, I don't feel it should be, should it be available to Discovery? No. Should you guys have it from Microsoft's perspective? No. you shouldn't collect this intently. I agree with that statement that in terms of, no. but I do understand that data makes difference to my life, and things change. As long as that company is open with the intent. I am willing to share it." (Respondent_JJ)*

Respondent_GS builds on Respondent_JJ's view and adds a philosophical layer of 'the greater good' to the concept of intent.

> *"I think if it's done for the greater good it's okay, I mean it becomes a very philosophical problem and I think that's part of the reason why there is this aggressive focus on building alternative technologies that can preserve and you can choose to have your stuff exposed or not I think. For me personally, if it is for the greater good that's okay, some of it's a bit irritating where you might get messages based on whatever and their trying to predict your behaviour and well it's there for the moment but it's pretty childish but I think there is going to be a maturity that comes into…" (Respondent_GS)*

Conclusion:

Possession of personal data is not the greatest concern. The intended use, however, has far reaching implications.

5.7 Summary of Findings

The point below summarises the relevant findings from the exploratory interview analysis.

Overarching Big Data and privacy concepts:

- Big Data represents Information Assets characterised by high volume, velocity and variety that require technological and analytical intervention (or rather veracity) for their transformation into value
- Not only do Big Data and business intelligence have a significant impact on the micro and macro-economic landscape of the business environment, they also progressively impact on various aspects of society at large
- Privacy is a human right and careful consideration should be accorded to the application of Big Data techniques to consumers and citizens lives. Ethical theories of egoism and utilitarianism complicate the right to privacy
- Criticism against Big Data is seen as a "necessary evil" as the rewards outweigh the risks. Additionally, the level of risk a customer is willing to take on is directly proportional to the value that will be derived from the service
- Although the legal premise of privacy legislation is sound, it has failed to keep pace with globalisation. With the relentless improvement and expansion of technological capabilities as well as the changing ways in which individuals

create, share and use personal data, privacy and in some ways, security legislation frameworks, have been unsuccessful in meeting their goal.

Personal Data Model:

- The premise of consumer empowerment with the individual as the centre of personal data collection, management and use, was unanimously welcomed and introduced the intellectual coherence and property base theory of personal data. However, its technical feasibility was questioned in as far as the system's ability to capture all-encompassing data – dynamic and static – was concerned.
- The concept of selective disclosure supports the ability of customers to share their data selectively, without disclosing more personal data than they wish to. However, data veracity is required to mitigate the data minimisation risk within the model.
- Personal data would lose its intellectual property coherence should the premise of control over purpose and duration not be included within the PDS model.
- The premise of signalling – "pull" action rather than "push" – was not deemed as a foundational block of the model, but raised questions of whether the effect of customisation would increase costs rather than achieve the desired cost reduction
- Identity management and security should be entrenched at a foundational layer of the PDS model. The introduction of Blockchain technology is a viable solution for various technical shortcomings of the model
- The ability to move personal data form one provider to another using standard data formats and interface protocols results in a moot point as the transfer of personal data is no longer required via the Blockchain platform
- Reliance on regulatory institutions only extends as far as the effectiveness of the institution. Conversely, government regulators and legislation should not take ownership of this PDS model; Blockchain technology provides for the ultimate accountability.

Additional concepts and considerations:

- The PDS, and organisations in general, assume users' and consumers' understanding of data and privacy.
- Possession of personal data is not the greatest concern. The intended use by those who have access to it, however, has far reaching implications.

CHAPTER 6: DISCUSSION OF RESULTS

6.1 Introduction

Chapter 6 relates findings established in Chapter 5 to pertinent literature reviewed in Chapter 2. This chapter associates the insights provided through the ten qualitative, exploratory interviews with Big Data and privacy subject matter experts (SMEs) with the appraised literature. The qualitative interview transcript analysis enabled the researcher to establish confirmation for or against each of the research propositions outlined in Chapter 3.

Following a review of the original framework developed, findings identified in Chapter 5 are presented in terms of the framework. Subsequent to this, each research proposition is reviewed in turn; finally, an adapted framework is presented.

6.2 Application of an *a priori* framework

The research pursued a line of investigation concerning the validity and feasibility of the guiding principles as well as the foundational elements of a Personal Data Store (PDS). Given the growing concern that companies and governments are not protecting, but rather exploiting, personal consumer data, as highlighted and described in the introduction to this paper, an *a priori* framework was developed based on the literature reviewed, encompassing eight primary elements to move individuals from an "Attention Economy' to 'Intention Economy'.

The framework was thereafter used to guide the development of the research propositions outlined in Chapter 3 as well as the interview schedule utilised and those results as detailed in Chapter 5.

For ease of reference the conceptual framework developed in Chapter 2 is re-presented in Figure 6.1 below.

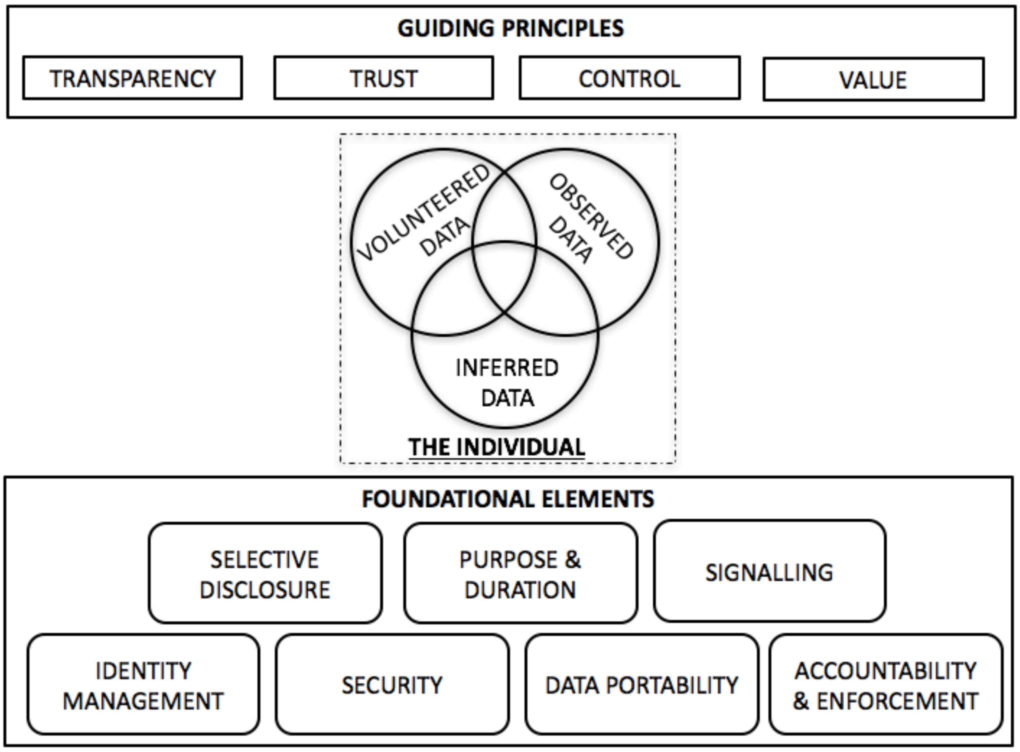


Figure 6-1 *A priori* Personal Data Store (PDS) framework

Source: Researcher's own construction

The vision of a PDS as presented at the World Economic Forum (2011) promises:

- Greater individual control over personal data, digital identity and online privacy, as well as increased compensation or shared value for providing others with access to personal data
- Disparate silos of personal data held in corporations and government agencies will more easily be exchanged to increase utility and trust among people, private firms and the public sector
- Governments' need to maintain stability, security and individual rights will be met in a more flexible, holistic and adaptive manner.

These guiding principles of transparency, control, trust and value also complement the 4 V's of Big Data (De Mauro, Greco & Grimaldi, 2015) whilst seeking to address the criticisms raised in terms of the application of Big Data (Tene, 2011).

User-centricity (the individual) is marked as the principal element or rather the focal point of a personal data store (PDS) ecosystem. Hinging user-centricity, this point aligns stakeholders' interests and realises the vision of the PDS (World Economic Forum, 2011).

A PDS includes the three types of personal data, i.e. volunteered, observed and inferred data, is enabled via data warehousing technology and utilises universal personal and mobile computing to assist individuals manage personal data as a personal asset (Rubenstein, 2013). The concept of user-centricity includes the right to access, update and rectify one's individual information (Tene, 2011).

Furthermore, outlined by Searls (2012), complemented by Tene and Polonetsky (2012; 2014), and formulated by Rubenstein (2013) a PDS encompasses eight foundational elements:

1. **User-centricity:** positioning an individual as the centre of data collection, management and use
2. The concept of **selective disclosure:** supports the ability of customers to share their data selectively, without disclosing more personal data than they wish to
3. **Control over the purpose and duration:** of primary and secondary uses of a user's personal data. This control may be achieved via an "owner data agreement" and/or by technical means such as Data Rights Management (DRM) of meta-data tagging
4. The basis of **signalling:** a means for individuals to express demand for goods or services in open markets, not tied to any single organisation, will promote higher quality advertisements as traditional targeting, capturing, acquiring, managing and locking a customer in would not be necessary. Customisation of products similarly empowers vendors to better tailor product / service propositions to a customer and promote revenue returns.
5. **Identity management:** entrenched in the process of interacting and transacting, organisations in both the private and public sector need to be confident that the person they are dealing with is who they say they are. Usually, this assurance is given by an agreed "gold standard" piece of identification such as a passport, or bank account. Personal Data Stores can help streamline these identity assurance processes by linking verifications to such data (World Economic Forum, 2011). Identity management within a PDS will manage tasks such as the authentication and use of multiple identifiers while preventing correlation, unless permitted by the user
6. **Security:** embedded within the technology platform and transaction layer
7. The proposition of **data portability:** includes a user's ability to move all of data from one provider to another using standard data formats and interface protocols. This flexibility allows for better competition and service delivery

8. **Accountability** of all stakeholders as well as **enforcement:** of protection and securing of personal data in accordance with the rights and permissions established by agreement and/or enforced by tagging mechanisms and enforcement under self-regulatory guidelines and legal mandates, both backed by comprehensive auditing.

From a technological, policy and sociological sense, all stakeholders need to embrace the construct and underlining framework. This requires innovation around user-centricity and trust. Additionally, defining global principles and strengthening the dialogue between regulation and the private sector will provide for a knowledge sharing bionetwork for using and sharing data (World Economic Forum, 2011).

In the following section the findings of Chapter 5 are related to the theory per each research proposition. For ease of reference each individual research proposition has been repeated at the start of each section.

6.3 Discussion of Research Proposition 1: Guiding Principles

From the literature review it is proposed that a PDS would require the guiding principles of transparency, control value and trust.

6.3.1 The four V's of Big Data

Central to research proposition 1, are the underlying theories of Big Data and business intelligence.

As discussed in the literature review, De Mauro, Greco and Grimaldi (2015, p.103) propose a consensual definition where *"Big Data represents Information assets characterised by High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value"* by looking at both the existing definitions of Big Data and at the main research topics associated with it:

- "Volume", "Velocity" and "Variety", describes the characteristics of Information involved
- Specific "Technology" and "Analytical Methods", to clarify the unique requirements strictly needed to make use of such Information; and
- Transformation into insights and consequent creation of economic "Value", as the principal way Big Data is impacting companies and society.

The results, as presented in the “Understanding of Big Data” (Section 5.6.1), discussed the contextual understanding of the overarching themes and supported the definition in so far as Big Data represents information assets characterised by high volume, velocity and variety that require technological and analytical intervention (or rather, veracity) for its transformation into value. Furthermore, merely the collection of static data without analytical methods for its transformation into “value” does not constitute Big Data in its truest form.

Likewise, the research findings outlined in the “Benefits of Big Data” (Section 5.6.2) stressed the positive attributes of Big Data and particularly in Machine Learning technology that powers many aspects of modern society. This includes anything from web searches to content filtering on social networks to recommendations on e-commerce websites (LeCun, Bengio & Hinton, 2015). In layman’s terms, Machine Learning enables an understanding of an individual’s context and has the power to highlight “golden threads” to identify and eliminate security incidents and events. Consensus amongst subjects is that organisations are not exploiting Big Data technologies enough.

As previously mentioned Dean, DiGrande, Field and Zwillenberg (2012) estimate that the Internet economy amounted to US$2.3 trillion in value in 2010, or 4.1% of total GDP, within the G20 group of nations and provides for a strong link between an effective data management strategy and financial performance.

Conventional machine-learning techniques were limited in their ability to process natural data in their raw form. However, deep learning (a class of systems’ learning techniques) is making major advances in solving problems that have resisted the best attempts of the artificial intelligence community. Deep learning is superior at discovering intricate structures in high-dimensional data and is therefore applicable to many domains of science, business and government (LeCun, Bengio & Hinton, 2015). For example, deep learning has outstripped other machine-learning techniques at predicting the activity of potential drug molecules (Ma, Sheridan, Liaw, Dahl & Svetnik, 2015) analysing particle accelerator data (Ciodaro, Deva, De Seixas, Damazio, 2012; Kaggle, 2014), reconstructing brain circuits (Helmstaedter, Briggman, Turaga, Jain, Seung, Denk, 2013) and predicting the effect of mutation in non-coding DNA on gene expression and disease (Leung, Xiong, Lee & Frey, 2014; Xiong, Alipanahi, Lee, Bretschneider, Merico, Yuen, Hua, Gueroussov, Najafabadi & Hughes, 2015).

Section 5.6.2 – the benefits of Big Data – concluded that the analytical techniques applicable to Big Data have had a significant impact on the micro and macro-economic landscape of the business and societal environment and are aligned with LeCun, Bengio and Hinton's (2015) prediction: that deep learning will have many more successes in the near future as it requires very little engineering by hand and can therefore take advantage of the developments in increased computational power and data. Advances in deep neural networks will just accelerate this progress.

6.3.2 Privacy as a human right

As mentioned earlier in this research, the privacy enigma has arisen to the level of a global debate primarily due to the revelations of Edward Snowden (Greenwald, MacAskill & Poitras, 2013). The worldview that no real harm comes from mass surveillance is grounded in the premise that there are two kinds of people in the world: good people and bad people. Greenwald (2014) argues the individuals that have this view are actually engaged in a very extreme act of self-deprecation. Personal privacy findings (Section 5.6) categorically viewed privacy as a human right and that careful consideration should be lent to the application of Big Data techniques to consumers' and citizens' lives.

This prompts the question: Is privacy in fact a human right?

In an investigation of western modernist theories, utilitarian moral philosophies presented an attempt to establish privacy as a universal value that is connected to important moral features such as dignity or well-being. According to Crane and Matten (2015, p101) utilitarianism argues that "an action is morally right if it results in the greatest amount of good for the greatest amount of people affected by the action."

In accordance with, *inter alia*, this moral philosophy the United Nations General Assembly adopted resolution 68/167 which expressed deep concern at the negative impact that surveillance and interception of communications may have on human rights. The General Assembly avowed that the rights held by people offline must also be protected online, and it called upon all States to respect and protect the right to privacy in digital communication (Peterson, 2008).

Pursuant to this, the European Union Court of Justice delivered a milestone ruling in the case of *Digital Rights Ireland*, the court declared the Data Retention Directive — an EU legislative act requiring telecommunications service providers to retain for up to two years all metadata from every EU citizens' emails, text messages, and telephone calls and to make these available to national security agencies for investigatory purposes – to be in violation of the rights to privacy and data protection enshrined in the European Union Charter of Fundamental Rights (Fabbrini, 2015).

However, in contrast to the utilitarianist position, objectivism offers an egoist approach to ethics that values individual privacy on rational, self-interest grounds (Drake, 2015). According to Crane and Matten (2015, p100) egoism follows the view that “an action is morally right if the decision-maker freely decides in order to pursue either their (short-term) desires or their (long-term) interests”. By applying objectivist principles to an organisational and societal context, we observe that citizens and governments should not violate fellow citizens’ privacy for short-term gains. Furthermore, Drake (2015 p.1) observes that “privacy can be protected without distinct rights to privacy. Rather, objectivism’s conception of rational self-interest suggests that long-term flourishing is the proper end of individuals and businesses, predicated on, amongst other things, respecting privacy and enforcing individual rights”.

A conundrum however exists in the case of crime and terrorism prevention. Should an individual’s right to privacy be denied where the “needs of the many outweigh the needs of the few”?

Fabbrini (2015) argues that in the case of the European Union Court of Justice in 2014, the court did not deny the importance of fighting crime and protecting national security. It advanced a strict proportionality framework, requiring that any interference with the broad understanding of privacy and data protection be strictly necessary to the attainment of the desired goal.

As the process of translating rights for online contexts deepens, conceptual, political and practical issues will continue to arise. There has been resistance to the idea of digital privacy from states involved in mass surveillance, which in itself points to privacy as having at least some rhetorical utility (Joice, 2015)

6.3.3 Unfavourable effect of Big Data qualified as a necessary evil

The harvesting of large sets of personal data and the use of cutting edge analytics fuel growing privacy concerns. Protecting privacy will become harder as information is multiplied and shared ever more widely among multiple parties around the world. As more information regarding individuals’ health, financials, location, electricity use and online activity percolates, concerns regarding profiling, tracking, discrimination, exclusion, government surveillance and loss of control arise (Daniel, 2006).

Noted criticisms against Big Data (Section 5.6.4) such as the incremental effect, behaviour modification and automated decision making, predictive analytics and access and exclusion were viewed as a “necessary evil” as the rewards outweigh the perceived risks.

Ohm (2010) describes this incremental effect as the “database of ruin”, chewing away, byte by byte, on an individual’s privacy until his or her profile is completely exposed. However, the findings concerning “criticisms against Big Data” (Section 5.6.4) highlighted that customers and consumers lacked the understanding that certain service benefits, such as identity management, are derived from the incremental effect.

The influence of Machine Learning, predictive analytics that influence online behavioural advertising and personalisation applications, raises concerns around discrimination, self-determination and the narrowing of choice (Tene, 2011). Conversely, certain desired social behaviours can also be stimulated by behaviour modification as seen in the influence of reward based incentives. Criticisms against Big Data findings (Section 5.6.4) provide practical examples in the case of “texting and driving”. Technologies are primed to modify this detrimental behaviour by utilising location based technologies and shutting down certain functionalities within a mobile phone until the vehicle has come to a full stop or alternatively, providing a “flashing” notice to deter drivers.

Predictive analytics has various societal benefits and positive implications for healthcare, specifically in the field of preventative care and early detection. Likewise, predictive analytics also has a useful application in law enforcement, national security, credit screening, insurance and employment. However, as noted by Tene and Polonetsky (2012) as well as Miller (2014) this raises an ethical dilemma where discrimination is prevalent in data profiling as an unfavourable consequence.

Thornhill’s (2016) article in the *Financial Mail*, argues that the United States are already utilising computer-generated risk scores to sentence convicted criminals and impose parole decisions. This statement was supported in the case of Eric Loomis, where he was awarded a six-year prison sentence; a length determined in part not just by Loomis’s criminal record, but also by his score on the COMPAS scale, an algorithmically determined assessment that aims, and claims, to predict an individual’s risk of recidivism (Garber, 2016).

This application of risk assessment algorithms revisits Orwell (1949) and Dick’s (1956) exploratory work on the concept of a predictive crime model. An excerpt from Dick’s *Minority Report* (1956, p 232) observes that underlying dangers exist where “...the commission of crime itself is absolute metaphysics. We claim they’re culpable. They, on the other hand, claim they’re innocent. And, in a sense, they are innocent.” He concludes: “In our society we have no major crimes. But we do have a detention camp full of would-be criminals.”

De Mauro, Greco and Grimaldi (2015) as well as Boyd and Crawford (2012) hypothesise that the split between information-rich and data-lacking companies may create a new digital divide that can slow down innovation in the sector. Specific policies will have to be promoted and data is likely to become a new dimension to consider within antitrust and non-competitive regulations.

Criticisms against Big Data findings (Section 5.6.4) recognised that the level of risk a customer is willing to take on is directly proportional to the value that will be derived from the service. This finding acknowledges that trade-offs are the natural realm of economics (Acquisti, 2010).

On the one hand, individuals want to protect the security of their data and avoid the misuse of information they pass to other entities. However, individuals also benefit from sharing with peers and third parties information that makes mutually satisfactory interactions possible. Organisations face that same trade-off. The entity wants to know more about the parties they interact with, tracking them across transactions. Yet, they do not want to alienate those parties with policies that may be deemed too aggressive. Acquisti (2010) argues that ultimately, the economic consequences of information sharing for all parties involved (the data subject and the actual or potential data holder) may be welfare enhancing or diminishing. In choosing the balance between sharing or hiding one’s personal information (and in choosing the balance between exploiting or protecting individuals’ data), both individuals and organisations face complex, sometimes intangible and often, ambiguous trade-offs.

6.3.4 Fallacy of privacy legislation

Current legislation models are constructed with the ultimate goal of permitting only legitimate procession of personal data. However, as with Big Data, the efficacy of the legislation is questioned and several shortcomings acknowledged.

The most prominent of shortcomings is the recurring dilemma of the law's inability to keep pace with technological change (Moses, 2007). This is not a new phenomenon: an early example of technology giving rise to legal problems concerns railroads, with topics ranging from property rights over track and eminent domain to liability for damages to employees, passengers, stock and land (Pierce, 1858).

The focus on technological change looks to changes in what is practically possible, rather than ordinary changes in behaviour or cultural practices. This excludes changes in social norms and customs that alter what we might be willing or wanting to do (Moses, 2007). Such changes are rarely so sudden and dramatic that the law's ability to keep pace is questioned. Moses (2007) maintains that where the law does respond to social change, it is rarely for the same reasons as it responds to technological change. Furthermore, not only will new technology frequently ground new law, it generates uncertainties regarding the application of existing law, observable in the early literature in areas as diverse as in-vitro fertilisation (Rozovsky, 1975), genetic testing (Hunderdwood & Cadle, 1996), computing processes (Blodgett, 1985) and nanotechnology (Fiedler & Reynolds, 1994).

Although the legal premise of privacy legislation is sound, it has failed to keep pace with globalisation. The findings for the efficacy of privacy legislation (Section 5.6.5) concluded that the relentless improvement and expansion of technological capabilities as well as the changing ways in which individuals create, share and use personal data, privacy and in certain ways, security legislation frameworks, have been unsuccessful in meeting their goal.

As with the ambiguity of PII, the concept of "Informed Consent" is fallible as individuals neither read nor understand privacy policies when the language is deemed ambiguous and time consuming. The efficacy of privacy legislation results (Section 5.6.5) illustrated that consumers blindly accept terms and raises questions concerning the level or perhaps type of education required to understand what one is consenting to.

Moreover, limiting privacy exposure through data minimisation is deemed as restricting the value and premise of Big Data. As suggested in the results for the efficacy of privacy legislation (Section 5.6.5), involving legislators in the research and innovation (R&D) cycle would stifle innovation.

In contradiction of Porter's (1991) hypothesis that social regulations not only induce innovation but also frequently enhance the competitiveness of the regulated firms, policy uncertainty does appear to precipitate both the negative and positive effects of expected future regulation, as indicated by Golec, Hegde and Vernon (2005), Taylor, Rubin and Hounshell (2005) and Aerni (2004). Nevertheless, classical theory holds that policy uncertainty causes businesses to delay investment decisions and the evidence presented here does not refute this. Most likely, the behaviour of firms operating under policy uncertainty depends upon the level of uncertainty and the profitability of the available actions given the range of expected regulatory alternatives. Higher uncertainty and larger differences in the expected profitability of innovation investments will tend to stifle innovation (Steward, 2010).

Rather than focusing on the need for technology-neutral legislation, one needs to consider how the legal system deals with dilemmas raised by technological change in a broader institutional context (Moses, 2007).

6.3.5 Conclusion

The research findings as well as the reviewed literature support the proposition that the aforesaid PDS would require the guiding principles of transparency, control value and trust. Therefor compiling the measurement scales of Big Data and privacy, economic theory holds firm. Although context dependent, Acquisti's (2010) economic theory of privacy professes that consumers are willing to trade privacy for convenience and economic value.

6.4 Discussion of Research Proposition 2: User-centricity

In terms of the literature review it is proposed that the central theme to PDS is the construct of user-centricity.

In support of Searls' (2012) intention economy, the World Economic Forum (2011), promotes end user-centricity as the pivot point in realising the vision of a personal data ecosystem. Likewise, Tene (2011) maintains that the right to access, update and rectify one's individual information remains distressingly underutilised. Few individuals are aware of their access rights and even fewer exercise them.

The PDS research (Section 5.6.6) was universally welcomed the premise of consumer empowerment and further introduced the notion of intellectual coherence and property base theory of personal data.

The monetary value of personal data is large and growing. Organisations view this information as a corporate asset and have invested heavily in software that facilitates the collection of consumer information, moving to profit from this trend (Schwartz, 2004). Moreover, a strong conception of personal data as a commodity is emerging and some individuals are already participating in the commodification of their personal data (Agre & Rottenberg, 1998). Conversely, under the concept of free alienability – the notion that an individual has the right to do what she wants with her personal information – propertisation of personal data might prevent restrictions from being placed on one's ability to trade personal data.

Schwartz (2004) theorised a model of commodification that encompasses five required elements:

i. Limitations on an individual's right to alienate personal information
ii. Default rules that force disclosure of the terms of trade
iii. Right of exit for participants in the market
iv. Damages to deter market abuses
v. Institutions to provide trading mechanisms to verify claims to propertised personal data (a verification function) as well as "police" compliance with bargained-for terms and legal safeguards (oversight function)

In contrast to criticisms, Rubenstein (2013) maintains that the principle of a personal data store (PDS) satisfies all elements in as far as that, firstly, PDS' are premised on use-transferability restrictions. Secondly the system not only combines use-transferability restrictions with an opt-in default but uses a tagging mechanism for enforcing initial and subsequent choices. Lastly, PDS's readily allow users to opt-in and -out of various services at any time, thereby enabling right of exit.

Schwartz (2004) identifies market-making, verification and oversight in a decentralised view; in terms of this theory, a higher market value for personal data might heighten our appreciation for it.

6.4.1 Conclusion

The research findings as well as reviewed literature endorse the proposition that the focal point of the PDS is the construct of user-centricity. The research further confirms that end user-centricity will provide greater control within the Big Data continuum.

6.5 Discussion of Research Proposition 3: Foundational Elements

Based on the findings from the literature review, it is proposed that for the validity and technical feasibility of the hypothesised PDS, all elements (selective disclosure, purpose and duration, signalling, identity management, security, data portability and accountability and enforcement) are required at a foundational level.

6.5.1 Selective disclosure

The concept of selective disclosure provisions the ability for users to share data selectively, without disclosing more personal data than they wish to.

The proposition specifies that organisations subscribe to updates from specific fields within the individual's PDS. To gain access, they have to sign the individual's terms and conditions; the individual can choose which organisation he or she wishes to accept or reject as a subscriber. Once the subscription is in place, every time the individual changes the relevant field in her/his data store, the subscribing organisation is alerted to this fact. Searls (2013) suggests that on the premise that perfect data is provided during the opt-in motion, operational costs associated with traditional data cleansing would not be necessary and will also remove the "guessing game" that data analytics play.

The results presented within the PDS (Section 5.6.6) supported the ability of customers to share their data selectively, without disclosing more personal data than they wish to. However, data veracity is required to mitigate the data minimisation risk within the model. Complementing the data veracity argument, respondents also questioned a system's ability to capture all-encompassing data; dynamic as well as static, and whether this is technically feasible. To this point Rubenstein (2012) argues that in order to be feasible technically, PDS's must meet two main requirements: security at a very high level and the ability to enforce privacy rights by "tagging" every unit of personal data with meta data describing privacy related requirements and preferences. Searls' (2013) suggestion of "perfect data" is provided during the opt-in motion: this would include "live" or "dynamic" data, mitigate the risk proposed by data minimisation and decrease traditional operational data cleansing costs.

6.5.2 Purpose and duration

Control over the purpose and duration of primary and secondary uses of a user's personal data. This control may be achieved via an "owner data agreement" and/or by technical means such as Data Rights Management (DRM) of meta-data tagging.

The results presented within the PDS (Section 5.6.6) supported this element as personal data would lose its intellectual property coherence, should the premise of control over purpose and duration not be included within the PDS model.

Exploring rights enforcement via tagging, Zittrain (2000) argued that digital rights management (DRM) systems offer the basis for privacy adoption. According to Zittrain (2000) trusted systems structure "rights" into a calculable framework that is enforced by technology. Complementing Zittrain's (2000) proposition, Korba and Kenny (2002) have since proposed a "privacy rights management" (PRM) system. Rubenstein (2012) argues that under this model, the data controller acts as the enforcer of the usage requirement for personal data and manages the collection, storage and processing of personal data from the data subjects.

6.5.3 Signalling

A means for individuals to express demand for goods or services in open markets, not tied to any single organisation. The basis of signalling will promote higher quality advertisements as traditional targeting, capturing, acquiring, managing and locking a customer in would not be necessary. Customisation of products similarly empowers vendors to better tailor product / service propositions to a customer and promote revenue returns.

Contrary to Rubenstein's (2012) and Searls' (2013) opinions, the premise of signalling – "pull" action rather than "push" – was not deemed to be a foundation block of the model in the results presented for the PDS (Section 5.6.6), but raised the question of whether the effect of customisation would increase costs rather than the desired opposite. In contrast, respondents supported the basis of signalling that promotes higher quality advertisements. The balance between the increase in revenue and higher operational costs requires further exploration and would need to be measured on a case by case basis.

6.5.4 Identity management and security

Identity management and security functionalities within a personal data store aim to manage tasks such as authentication, access and exploitation with the use of multiple identifiers while preventing correlation, unless permitted by the user.

The results presented for the PDS (Section 5.6.6) universally found that, without security and identity management as a foundational element, the hypothesis of the PDS would not succeed. Furthermore, respondents introduced the fundamentals of Blockchain technology as a viable solution for various technical, identification and security shortcomings of the model.

Within the global financial industry, Bitcoin has proved that trusted, auditable computing is possible using a decentralised network of peers accompanied by a public ledger. The distributed public ledger records transactions of things of value. In its rawest form, Blockchain is an automated way to record all transactions in a way that promotes trust and decreases cost. If this were coupled with distributed nodes that could be globally housed, this makes it practically impossible to corrupt this method of record keeping or transaction logging. The work developed by Zyskind, Nathan and Pentland (2015) demonstrates that by combining Blockchain and off-Blockchain storage, a personal data management platform – focussed on privacy – could be constructed.

Zyskind, Nathan and Pentland (2015) recognised the related work by de Montjoye, Shmueli, Wang, & Pentland (2014) as well as Rubenstein (2013) and Searls (2014), illustrating the model for autonomous deployment of an open PDS with mechanisms for returning computations on data, thus returning answers instead of the raw data itself. However, the recent increase in reported incidents of surveillance and security breaches, compromising users' privacy, calls the current model into question.

By implementing protocols that turns a Blockchain into an automated access-control manager, Zyskind, Nathan and Pentland's (2015) protocol provides a solution that does not require trust in a third party. Unlike Bitcoin, transactions in the system are not strictly financial, but rather used to carry instructions, such as storing, querying and sharing data. Figure 6.2 provides an overview of the system.

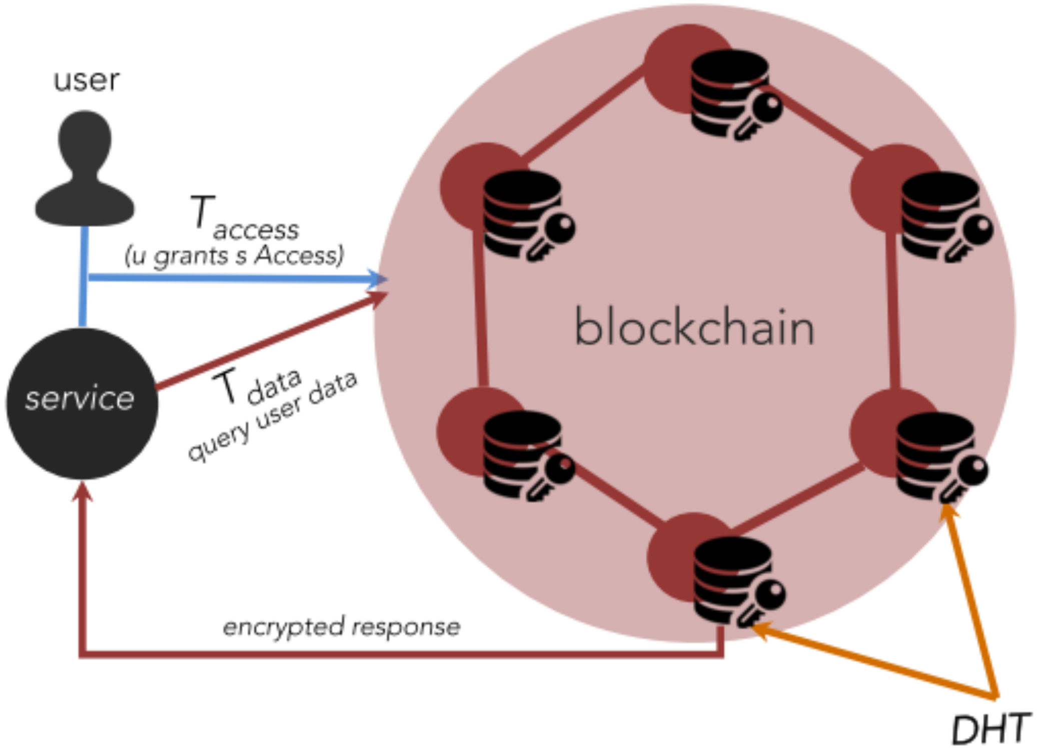


Figure 6-2: Overview of decentralised platform

(Zyskind, Nathan & Pentland, 2015)

As illustrated in Figure 6.2, Zyskind, Nathan, Pentland's (2015) system includes three entities: mobile phone users interested in downloading and using applications; services, the providers of such applications who require processing personal data for operational and business-related reasons, such as targeted ads and personalised service and nodes: entities entrusted with maintaining the Blockchain and a distributed private key-value data store in return for incentives. While users in the system normally remain (pseudo) anonymous, the system aims to store service profiles on the Blockchain and verify their identity. Distributed privacy via Blockchain technologies and related technologies of Bitcoin 2.0 is ground-breaking and requires further study.

Personal data, and sensitive data in general, should not be trusted in the hands of third-parties, where they are susceptible to attacks and misuse. In solidarity with Searls (2012), Rubenstein (2013) and Tene and Polonetsky (2014), Zyskind, Nathan and Pentland (2015) promote the notion that users should own and control their data without compromising security or limiting companies' and authorities' ability to provide personalised services. Furthermore, using a decentralised platform, making legal and regulatory decisions about collecting, storing and sharing sensitive data should be simpler.

The validity and feasibility of the *a priori* framework requires identity management and security entrenched at a foundational layer of the PDS model.

6.5.5 Data portability

This proposition includes a user's ability to move all data from one provider to another using standard data formats and interface protocols. This flexibility allows for better competition and service delivery.

The results presented within the PDS (Section 5.6.6) found that the ability to move personal data from one provider to another using standard data formats and interface protocols is a moot point as the transfer of personal data is no longer required via the Blockchain technology platform.

6.5.6 Accountability and enforcement

Accountability of all stakeholders as well as enforcement of protection and securing of personal data in accordance with the rights and permissions established by agreement and/or enforced by tagging mechanisms and enforcement under self-regulatory guidelines and legal mandates, both backed by comprehensive auditing, is necessary.

Reliance on regulatory institutions goes just as far as the effectiveness of the institution. Conversely, the results presented within the PDS (Section 5.6.6) remain distrustful as regards the abilities of regulators and propose that government regulators and legislation should not take ownership of this PDS model, arguing that Blockchain technology provides for ultimate accountability. Moreover, Zyskind, Nathan and Pentland (2015) postulate that laws and regulations could be programmed into the Blockchain itself, so that they are automatically enforced. In other situations, the ledger could act as legal evidence for accessing (or storing) data, since it is (computationally) tamper-proof.

6.5.7 Conclusion

In contradiction to the reviewed literature, the research findings led to the conclusion that the validity and feasibility of the *a priori* framework requires identity management and security entrenched at a foundational layer of the PDS model. Furthermore, the research findings argue that selective disclosure, purpose and duration, signalling and data portability are seen as additional value added elements within the framework, but are not required at a foundational level.

6.6 Additional Constructs Identified

6.6.1 Education

Throughout Chapter 5 as well as the literature reviewed, the PDS (and organisations in general) assume that users and consumers understand the intricacies of data, the processing of data and the related privacy constructs. The opposite is however true.

Literature recognises that the social and economic impact of technology is widespread and accelerating and predicts that 90% of the entire global population will be connected to the internet within the next 10 years. Park's (2016) summation in the World Economic Forum article: "8 digital skills we must teach our children", put children at the centre of this dynamic change.

Park (2016) frames the digital age gap and argues that the way children use technology is very different from adults. This gap makes it difficult for parents and educators to fully understand the risks and threats that children face online. As a result, adults may feel unable to advise children on the safe and responsible use of digital technologies. Likewise, this gap gives rise to different perspectives on what is considered acceptable behaviour. Park (2016) further argues that the problem lies in the fast and ever evolving nature of the digital world, where proper internet governance and policies for child protection are slow to catch up, rendering them ineffective.

To deepen the conversation around digital education, to consider a new form of inequality – "The Digital Divide" or "Digital Inequality" – is needed. Ali (2011, p?) provides the definition of the global digital divide as the unequal distribution of information and communication technology across nations, commonly described as the "gap between the information haves and have-nots". Ali (2011) further recognises that within academic circles it is well established that the digital divide encompasses more than physical access to information systems, but that it is also a function of how these systems are used. Tsatsou (2011) complements Ali (2011) and provides the argument that the web of cultural traits in a society, with its own gaps and disparities, as well as policy and regulation dynamics, are in a constant dialogue with technology, together influencing social inclusion (or exclusion) and participation.

The research findings conclude that the assumption of education or digital intelligence is required at a foundational layer within the PDS. Although the construct of the contribution of gender, age, education level of skill or understanding, internet experience and the amount of internet use, has been investigated, this facet of digital privacy education, not only within mature markets but also in developing nations, remains relatively unexplored.

The World Economic Forum in conjunction with Park (2016) provide a preliminary framework for digital intelligence (DQ) encompassing digital rights, literacy, communication, emotional intelligence, security, safety, use and identity.

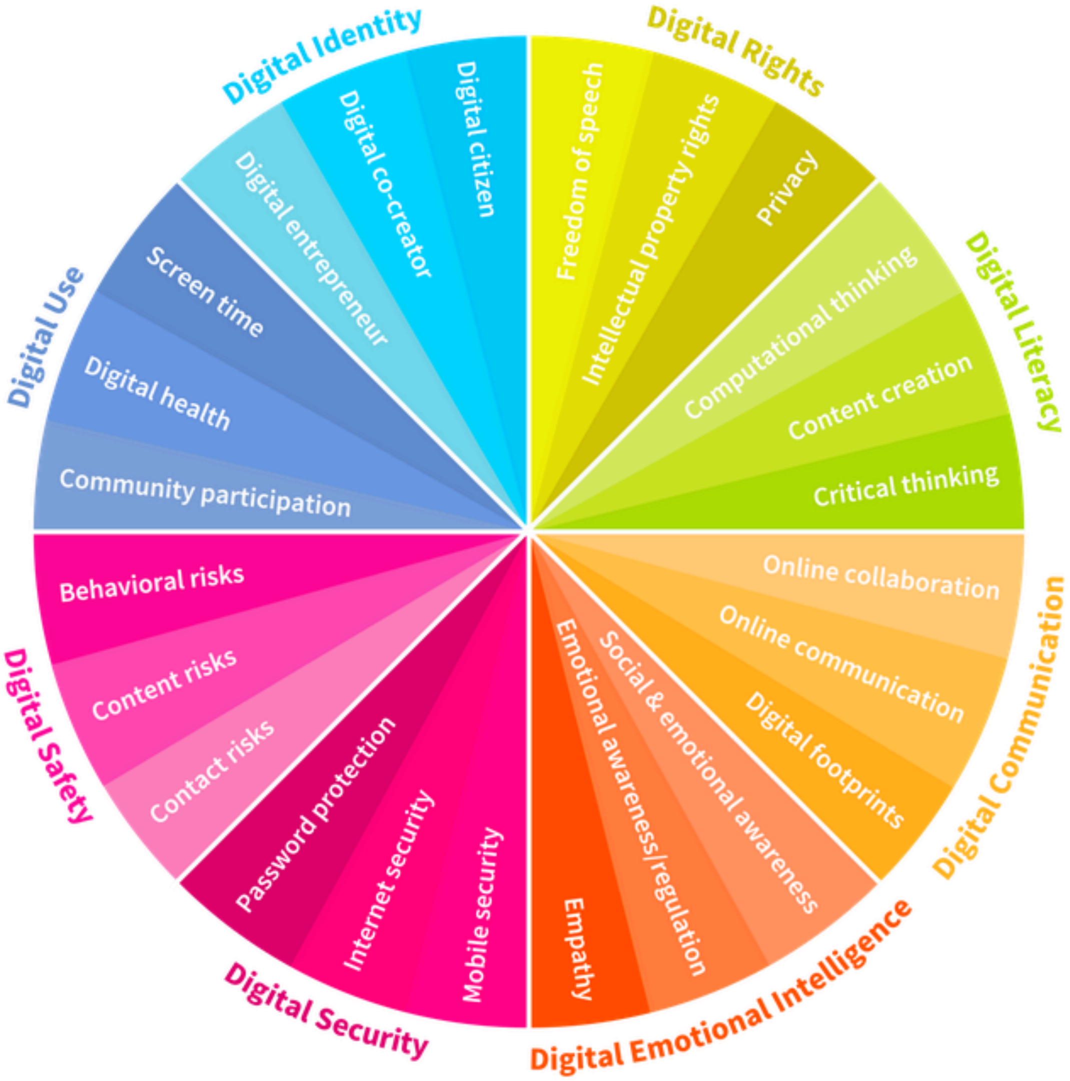


Figure 6-3: Digital intelligence infographic

(Park, 2016)

6.6.2 Intent

As with the educational construct, the overarching theme of "Intent" continues to resurface. Possession of personal data is not the greatest concern. The intended use, however, has far-reaching implications.

In alignment with the exploration of privacy as a human right, the construct of intent also lends itself to the ethical or moral philosophies. Utilitarian moral philosophies present a case for the "greater good" whereas objectivism, or rather egoism, advocates for the rights of the individual.

During an interview Respondent_JJ phrased it best: "… from a philosophical point of view, there is no right answer and there is only my point of view."

The research findings conclude that at this point in time, the guiding principles of the PDS (transparency, control, trust and value) are sufficient to ensure the validity and feasibility of the said PDS. However, the results also suggest that the literature around Big Data privacy and intent requires further investigation.

6.7 Updated *A Priori* Framework

Considering the findings of the research it became clear that the *a priori* framework required adaptation. Figure 6-4 below illustrates the adapted *a priori* framework of a PDS.

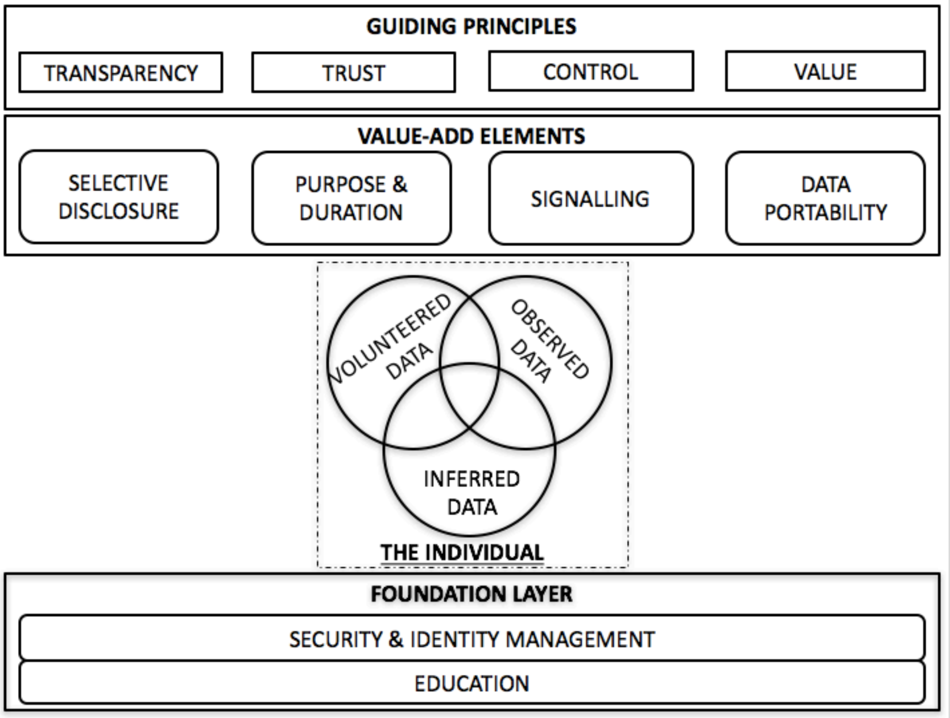


Figure 6-4: Adapted *a priori* framework of a Personal Data Store (PDS)

Source: Researcher's own construction

Research results conclude that the validity and feasibility of the *a priori* framework requires that digital intelligence (DQ) is assumed at the foundation layer of the framework. Additionally, the research findings and literature support the premise of user-centricity as the focal point within the PDS and conclude that end user-centricity will provide greater control within the Big Data continuum.

In contraction of the reviewed literature, the research findings conclude that the validity and feasibility of the *a priori* framework requires identity management and security entrenched at a foundational layer of the personal data store model. Furthermore, the research findings maintain that selective disclosure, purpose and duration, signally and data portability are regarded as additional value added elements within the framework but are not required at a foundational level.

At this point in time, the guiding principles of the personal data store (transparency, control, trust and value) are adequate to ensure the validity and feasibility of the PDS. However, the results also suggest that the literature around Big Data “privacy” and “intent” requires further investigation.

CHAPTER 7: CONCLUSION

7.1 Introduction

In this chapter, the findings of Chapter 5 and the discussions in relation to the theory in Chapter 6 are consolidated and the initial *a priori* framework adapted according to the outcomes of this research. This is followed by a discussion of the implications of the framework for management, limitations of this research and suggestions for possible avenues of future research.

7.2 Principal Findings

7.2.1 Summary of the finding of this research

The research sought to investigate the validity and feasibility of the guiding principles as well as the foundational elements of a Personal Data Store (PDS) as outlined by Searls (2012) complemented by Tene and Polonetsky (2012; 2014) and formulated by Rubenstein (2013). Given the growing concern that companies and governments are not protecting, but rather exploiting, personal consumer data, an *a priori* framework was developed encompassing eight primary elements to move individuals from an “Attention Economy” to an “Intention Economy”.

The guiding principles of the PDS (transparency, control, trust and value) ensure its validity and feasibility whilst complementing the 4 V’s of Big Data (De Mauro, Greco & Grimaldi, 2015). Furthermore, in support of Searls’ (2012) intention economy, user-centricity will provide greater control within the Big Data continuum by acting as the pivot point in realising the vision of a personal data ecosystem.

However, because personal data, and sensitive data in general, should not be trusted in the hands of third parties, where they are susceptible to attacks and misuse, identity management and security must be entrenched at a foundational level of the model.

In theoretical realms, the assumption of universal skill and understanding, or rather of digital intelligence, is sound. Conversely, in practice “The Digital Divide” or “Digital Inequality” is a far greater reality. Therefore, digital intelligence is also required at a foundational layer within the framework of a PDS.

Value added attributes, such as selective disclosure, purpose and duration, signally and data portability allow for the commodification of personal data (Schwarz, 2004). Likewise, a decentralised view of personal data supports the conception of personal data as a commodity (Agre & Rottenberg, 1998).

Figure 7.1 re-presents the adapted *a priori* framework based on the findings of this research.

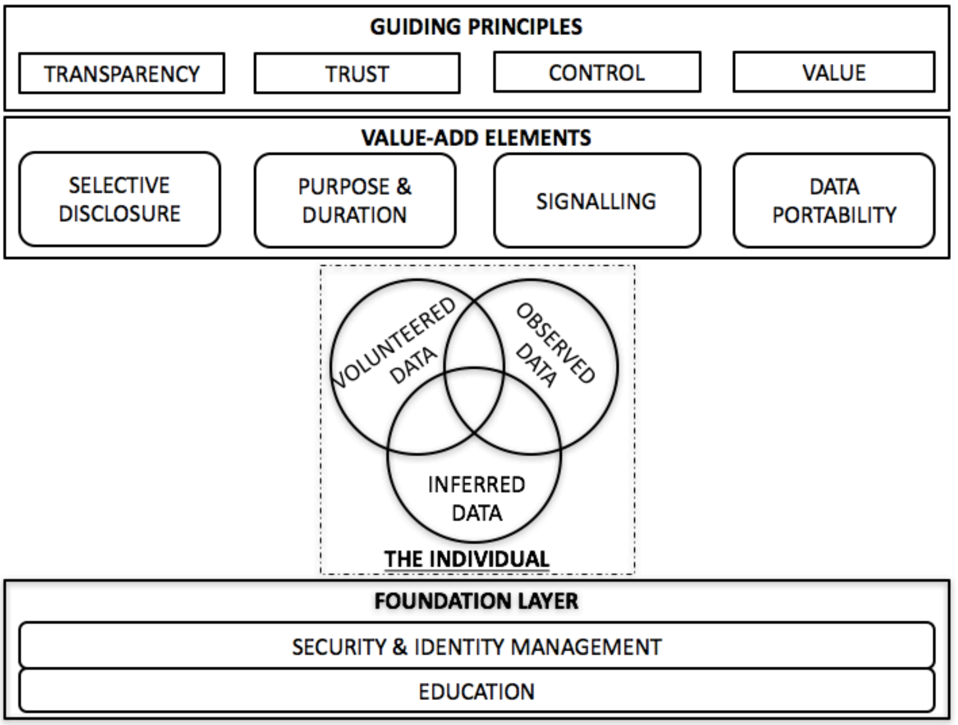


Figure 7-1: Adapted *a priori* framework of a Personal Data Store (PDS)

It is also hoped that this framework will provide a basis upon which further research into the implementation of a PDS can be built.

7.2.2 Contributions to Literature

From the discussions of the findings of this research regarding the literature already presented as part of Chapter 6, the following primary contributions to the literature on Big Data privacy continuum have been identified:

- In contradiction to Rubenstein (2013) and Searls (2012), these research findings conclude that without security and identity management as a foundational element, the hypothesis of the PDS will not succeed.
- The model proposed by Searls (2012), Tene and Polonetsky (2012; 2014), and Rubenstein (2013) is constructed on the assumption of skill and understanding within the universe of personal data and privacy. In practice “The Digital Divide” or “Digital Inequality” is a far greater reality. Therefore, digital intelligence is required at a foundational layer within the model.

7.2.3 Implications for Management

Similar to the revolutions that preceded it, the Fourth Industrial Revolution has the potential to raise global income levels and improve the quality of life for populations around the world. Schwab (2016) argues that, to date, those who have gained the most from it have been consumers able to afford and access the digital world; technology has made possible new products and services that increase the efficiency and pleasure of personal lives. Organisations have also benefited immensely, as technological innovation leads to a supply-side “miracle”, with long-term gains in efficiency and productivity.

However, given the growing concern and candid distrust in companies’ and governments’ ability to protect and not exploit personal consumer data, managers need to consider the following:

Firstly, the growing demand for transparency, security and consumer engagement will force companies to adapt the way they design, market and deliver products and services (Schwab, 2016). This, coupled with the growing monetary value of personal data (Schwartz, 2004), requires that organisations view this information as a corporate asset and invest heavily in the transparency, identity management and security of personal data (Rubenstein, 2013). Failure to do so could negatively impact organisational brand equity as well as customer loyalty.

Furthermore, the underlying theme of technological disruption and acceleration of innovation is evident. Even for the best connected and most well informed manager, the velocity of disruption is hard to comprehend or anticipate (Schwab, 2016). In the age of the Internet of Things (IoT) organisations churn out increasing volumes of transactional data, capturing trillions of bytes of information about their customers, suppliers and operations (Tene & Polonetsky, 2012). However, amplifying the rate of technological disruption yet failing to provide safe spaces where individuals can think free, divergent and creative thoughts will significantly diminish the progress organisations (and society) can enjoy.

Lastly, many industries are seeing the introduction of new technologies that create entirely innovative ways of serving existing needs and significantly disrupt existing industry value chains (Schwab, 2016). This is most evident within the global financial industry where Bitcoin has proved that trusted, auditable computing is possible using a decentralised network of peers, accompanied by a public ledger.

Disruption is also flowing from agile, innovative competitors who, thanks to access to global digital platforms for research, development, marketing, sales and distribution – illustrated by the work developed by Zyskind, Nathan and Pentland (2015) – can oust well-established incumbents faster than ever by improving the quality, speed, or price at which value is delivered (Schwab, 2016).

7.3 Limitations of the Research

7.3.1 Researcher bias

Researcher bias includes any factor which induces bias in the researcher's recording of responses; because exploratory research is quite subjective it is influenced by her or his perspectives. It is therefore important for the researcher to acknowledge those potential biases, as their context will have an influence on how she or he interprets the findings of the research (Creswell, 2014, p. 188; Saunders & Lewis, 2012), and the researcher's culture may create as much "blindness as insight" (McCracken, 1988, p. 6). Consequently, it must be recognised that the researcher has extensive experience working in Risk and Compliance as well as a broad understanding of the Information Technology industry. As a result, this may have biased some of the answers given by the respondent or may have placed too much emphasis on a particular theme.

7.3.2 Sampling bias

Yin (2009) notes that selecting new data collection units or interviewees as an offshoot of existing ones could be acceptable if the snowballing is purposeful, and not performed out of convenience. To avoid the pitfall of convenience bias, the reason for selecting units or interviewees must be defined and critiqued prior to the interview. Yin (2009) recommends distinguishing between a purposive reason and a merely convenient one.

The use of snowball sampling resulted in identifying a number of business partners that were closely involved in the researcher's multinational organisation as well as involved in some manner in the information technology industry. Three of the ten respondents were within the same multinational organisation as the researcher – while a further four were in close business partnership with the organisation. This may influence the transferability construct of the research and limit it to the said industry.

7.3.3 Respondent bias

All the respondents or participants exhibited a high level of comfort with the concepts and themes within the research. All but two of the respondents either actively use business intelligence models or lead teams that employ Big Data within the organisation. The remaining two respondents practiced within the current South African and European data privacy legislation framework.

While this may point to the fact that an understanding of technology as well as of legislative considerations related to privacy is an important aspect of an SME's understanding of business intelligence, no individuals nor representatives of companies were interviewed that did not have a clear understanding of Big Data in practice; hence no data were available to provide a valid counterpoint. This touches on the transferability of the research findings.

7.4 Recommendations for Future Research

Having validated the *a priori* framework of a PDS, the next step would be to test the avenues for implementation and adoption within relevant markets. In support of a decentralised personal data management platform developed by Zyskind, Nathan and Pentland (2015) the research contemplates combining blockchain and off-blockchain storage. The technical feasibility of distributed privacy via blockchain technologies and related technologies of Bitcoin 2.0 requires further analysis.

Moreover, the research identified three additional constructs that require further examination.

Firstly, in contradiction of Rubenstein (2012) and Searls (2013) the premise of signalling – a "pull" action rather than a "push" one – was not deemed as a foundation block of the model of a PDS but, instead, the researcher questioned whether the effect of customisation would increase costs rather than reduce them as desired. In contrast, the research supported the basis of signalling that promotes higher quality advertisements. This begs the question whether the balance between the increased revenue and higher operational costs should be measured on a case by case basis.

Secondly, although the construct of the contribution of gender, age, education level of skill and understanding, internet experience and the amount of internet use has been investigated, the facet of digital privacy education, not only within mature markets but also developing nations, remains relatively unexplored.

Finally, during an interview, Respondent_JJ declared that: "… from a philosophical point of view, there is no right answer and there is only my point of view." As with the educational construct, the overarching theme of "Intent" as well as the moral philosophies of utilitarianism and egoism continue to resurface and will continue to require further contemplation.

7.5 Conclusion

At the start of this research, in Chapter 1, it was recognised that technological innovation, efficiency and productivity come at the price of personal privacy and the researcher raised the question whether a balance could be struck between the competing principles of Big Data and an individual's right to privacy.

The findings of this research examined the validity and feasibly of the *a priori* framework, concluding that a balance is indeed possible. In matching the guiding principles of transparency, control, trust and value with a pivotal point of user-centricity as well as the foundational elements of identity management, security and digital intelligence, the model proposed by Searls (2012) and Rubenstein (2013) is not only valid but also feasible in addressing the Big Data/privacy conundrum.

Schwab (2016) recognises that change is needed at every level of society, from the individual and the personal to the institutional to the global, in taking responsibility for how we adapt to these technological changes and more importantly, challenges in what it means to be human, what it means to work, what it means to be completely embedded in this world – privately and publicly.

REFERENCES

Acquisti, A., & Gross, R. (2006). Imagined Communities: Awareness, Information sharing, and privacy on Facebook. *Privacy Enhancing Technologies*, 36–58. http://doi.org/10.1007/11957454_3

Acquisti, A., Lane, J., Stodden, V., Bender, S., & Nissenbaum, H. (2010). The economics and behavioral economics of privacy. In *Privacy, Big Data, and the Public Good* (pp. 76–95). New York, NY: Cambridge University Press. http://doi.org/10.1017/CBO9781107590205.005

Adomavicius, G., & Tuzhilin, A. (2005). Personalization technologies. *Communications of the ACM*, *48*(10), 83–90. http://doi.org/10.1145/1089107.1089109

Aerni, P. (2004). Risk, regulation and innovation: The case of aquaculture and transgenic fish. *Aquatic Sciences*, *66*(3), 327–341. http://doi.org/10.1007/s00027-004-0715-8

Agre, P. E., & Rotenberg, M. (1998). *Technology and privacy: The new landscape*. Cambridge, MA: MIT Press.

Ali, A. H. (2011). The power of social media in developing nations: New tools for closing the global digital divide and beyond. *Harvard Human Rights Journal*, *24*(2), 185–219.

Andrew Crane, & Matten, D. (2015). *Business ethics: Managing corporate citizenship and sustainability in the age of globalization*. Oxford: Oxford University Press.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. Retrieved October 31, 2016, from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Atzori, L., Iera, A., & Morabito, G. (2010). The Internet of Things: A survey. *Computer Networks*, *54*(15), 2787–2805. http://doi.org/10.1016/j.comnet.2010.05.010

Barnes, M. E. (2006). Falling short of the mark: The United States' response to the European Union's data privacy directive. *Northwestern Journal International Law and Business*, *27*, 171.

Barnes, S. B. (2006). A privacy paradox: Social networking in the United States. *First Monday*, *11*(9). http://doi.org/10.5210/fm.v11i9.1394

Barnett, J. M. (2002). Focus groups tips for beginners. *Texas Centre for the Advancement of Literacy & Learning, Occasional Research Paper*, *1*.

Barrett, B. (2016). You should go check Facebook's new privacy settings. Retrieved October 10, 2016, from https://www.wired.com/2016/06/go-check-facebooks-new-privacy-settings/

Bergström, A. (2015). Online privacy concerns: A broad approach to understanding the concerns of different groups for different uses. *Computers in Human Behaviour*, *53*, 419–426. http://doi.org/10.1016/j.chb.2015.07.025

Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Systems*, *50*(3), 602–613. http://doi.org/10.1016/j.dss.2010.08.008

Blodgett, N. (1984). Computer law quicksand: Pioneers in burgeoning field have little to guide them. *ABA Journal*, *70*(11), 32–33.

Bodhani, A., & Hayes, J. (2013). Cyber security: Small firms under fire. *Engineering & Technology*, *8*(6), 80–83. http://doi.org/10.1049/et.2013.0614

Boyd, D., & Crawford, K. (2012). Critical questions for Big Data. *Information, Communication & Society*, *15*(5), 662–679. http://doi.org/10.1080/1369118X.2012.678878

Bryman, A. (2008). Why do researchers integrate/combine/mesh/blend/mix/merge/fuse quantitative and qualitative research? *Advances in Mixed Methods Research*, 87–100.

Buhl, H. U., Röglinger, M., Moser, F., & Heidemann, J. (2013). Big Data. *Business & Information Systems Engineering*, *5*(2), 65–69. http://doi.org/10.1007/s12599-013-0249-5

Bunn, A. (2015). The curious case of the right to be forgotten. *Computer Law & Security Review*, *31*(3), 336–350. http://doi.org/10.1016/j.clsr.2015.03.006

Bygrave, L. A. (2001). Automated profiling - Minding the machine: Article 15 of the EC data protection directive and automated profiling. *Computer Law and Security Report*, *17*(1), 17–24. http://doi.org/10.1016/S0267-3649(01)00104-2

Capistrano, E. P. S., & Chen, J. V. (2015). Information privacy policies: The effects of policy characteristics and online experience. *Computer Standards & Interfaces*, *42*, 24–31. http://doi.org/10.1016/j.csi.2015.04.001

Cassell, C., & Symon, G. (2004). *Essential guide to qualitative methods in organizational research* (Vol. 2). New York, NY: Sage.

Cate, F. H. (2006). The failure of fair information practice principles. *Consumer Protection in the Age of the Information Economy*.

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From Big Data to big Impact. *MIS Quarterly*, *36*(4), 1165–1188.

Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A Survey. *Mobile Networks and Applications*, *19*(2), 171–209. http://doi.org/10.1007/s11036-013-0489-0

Ciodaro, T., Deva, D., de Seixas, J. M., & Damazio, D. (2012). Online particle detection with Neural Networks based on topological calorimetry information. *Journal of Physics: Conference Series*, *368*(1), 12030. http://doi.org/10.1088/1742-6596/368/1/012030

Clarke, R. (2011). An evaluation of privacy impact assessment guidance documents. *International Data Privacy Law*, *1*(2), 111–120.

Cook, T. (2002). Archives and privacy in a wired world: the impact of the personal information act (Bill C-6) on archives. *Archivaria*, *1*(53).

Correa Bahnsen, A., Aouada, D., Stojanovic, A., & Ottersten, B. (2016). Feature engineering strategies for credit card fraud detection. *Expert Systems with Applications*, *51*, 134–142. http://doi.org/10.1016/j.eswa.2015.12.030

Creswell, J. W. (2014). *A concise introduction to mixed methods research*. New York, NY: Sage.

Creswell, J. W., & Miller, D. L. (2000). Determining Validity in qualitative inquiry. *Theory Into Practice*, *39*(3), 124–130. http://doi.org/10.1207/s15430421tip3903_2

Cukier, K., & Mayer-Schönberger, V. (2013). The rise of Big Data: How it's changing the way we think about the world. *Foreign Affairs*, *92*(3), 28–40.

De Mauro, A., Greco, M., & Grimaldi, M. (2015). What is Big Data? A consensual definition and a review of key research topics. In *Proceedings of the 4th International Conference on Integrated Information* (Vol. 1644, pp. 97–104). http://doi.org/10.1063/1.4907823

De Wolf, R., Willaert, K., & Pierson, J. (2014). Managing privacy boundaries together: Exploring individual and group privacy management strategies in Facebook. *Computers in Human Behavior*, *35*, 444–454. http://doi.org/10.1016/j.chb.2014.03.010

Dean, D., DiGrande, S., Field, D., & Zwillenberg, P. (2012). *The connected world - the digital manifesto: How companies and countries can win in the digital economy*. *Boston Consulting Group* (Vol. 27).

Dick, P. K. (2002). *Selected short stories of Philip K. Dick*. New York, NY: Pantheon.

Donaldson, T., & Dunfee, T. W. (1994). Toward a unified conception of business ethics: Integrative social contracts theory. *Academy of Management Review*, *19*(2), 252–284. http://doi.org/10.5465/AMR.1994.9410210749

Donaldson, T., & Preston, L. E. (1995). The stakeholder theory of the corporation: Concepts, evidence, and implications. *Academy of Management Review*, *20*(1), 65–91. http://doi.org/10.5465/AMR.1995.9503271992

Drake, J. R. (2015). Asking for Facebook logins: An egoist case for privacy. *Journal of Business Ethics*, 1–13. http://doi.org/10.1007/s10551-015-2586-4

Duhigg, C. (2012). How companies learn your secrets. Retrieved May 8, 2016, from http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?_r=1

Duri, S., Elliott, J., Gruteser, M., Liu, X., Moskowitz, P., Perez, R., … Tang, J.-M. (2004). Data protection and data sharing in telematics. *Mobile Networks and Applications*, *9*(6), 693–701.

Epstein, R. (2015). How Google could rig the 2016 election. Retrieved October 31, 2016, from http://www.politico.com/magazine/story/2015/08/how-google-could-rig-the-2016-election-121548

Epstein, R., & Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences*, *112*(33), E4512–E4521. http://doi.org/10.1073/pnas.1419828112

Estrin, D., Culler, D., Pister, K., & Sukhatme, D. (2002). Connecting the physical world with pervasive networks. *IEEE Pervasive Computing*, *1*(1), 59–69.

Eurobarometer, F. (2008). *Data protection in the European Union citizens' perceptions*. Retrieved from http://ec.europa.eu/public_opinion/flash/fl_225_en.pdf

Euromonitor International. (2015). *Internet Users: Euromonitor International from International Telecommunications Union/OECD/national statistics*. Retrieved from http://0-www.portal.euromonitor.com.innopac.up.ac.za/portal/statistics/tab

European Central Bank. (2014). *Annual Report 2014*. Retrieved from http://www.ecb.europa.eu/pub/pdf/annrep/ar2014en.pdf

European Parliament, & European Commission. (1995). Directive 1995/46/EC on protection of individuals with regard to the processing of personal data on the free movement of such data. *Official Journal of the European Communities*. http://doi.org/ISSN 0378-6978

Evans, D. (2011). The Internet of Things: How the next evolution of the internet is changing everything. *CISCO White Paper*, (April), 1–11. http://doi.org/10.1109/IEEESTD.2007.373646

Fabbrini, F. (2015). Human rights in the digital age: The European Court of Justice ruling in the data retention case and its lessons for privacy and surveillance in the united states. *Harvard Human Rights Journal*, *28*(1), 65–95.

Fiedler, F. A., & Reynolds, G. H. (1993). Legal problems of nanotechnology: An overview. *Southern California Interdisciplinary Law Journal*, *3*, 593.

Foucault, M. (2007). *Abnormal: Lectures at the Collège de France, 1974-1975* (Vol. 2). London: Macmillan.

Foucault, M., Bertani, M., Fontana, A., Ewald, F., & Macey, D. (2003). *"Society Must Be Defended: Lectures at the Collège de France, 1975-1976* (Reprint, Vol. 3). London: Picador.

Friese, S. (2014). *Qualitative data analysis with ATLAS. ti*. New York, NY: Sage.

Gantz, J. F., & Chute, C. (2008). *The diverse and exploding digital universe: An updated forecast of worldwide information growth through 2011*. Retrieved from https://www.ifap.ru/library/book268.pdf

Gantz, J., & Reinsel, D. (2010). *The digital universe decade: Are you ready? IDC*. Retrieved from https://www.emc.com/collateral/analyst-reports/idc-digital-universe-are-you-ready.pdf

Garber, M. (2016). The Eric Loomis case and predictive crime assessments: When algorithms take the stand. Retrieved October 31, 2016, from http://www.theatlantic.com/technology/archive/2016/06/when-algorithms-take-the-stand/489566/

Gartner. (2014). Gartner says the Internet of Things will transform the data center. Retrieved October 10, 2016, from http://www.gartner.com/newsroom/id/2684616

Golec, J., Hegde, S., & Vernon, J. A. (2010). Pharmaceutical R&D Spending and threats of price regulation. *Journal of Financial and Quantitative Analysis*, *45*(1), 239. http://doi.org/10.1017/S0022109009990512

Golle, P. (2006). Revisiting the uniqueness of simple demographics in the US population. In *Proceedings of the 5th ACM workshop on Privacy in electronic society - WPES '06* (p. 77). New York, NY: ACM Press. http://doi.org/10.1145/1179601.1179615

Greenwald, G. (2014). *No place to hide: Edward Snowden, the NSA, and the US surveillance state*. London: Penguin.

Greenwald, G., MacAskill, E., & Poitras, L. (2013). Edward Snowden: The whistle-blower behind the NSA surveillance revelations. Retrieved October 31, 2016, from https://www.theguardian.com/world/2013/jun/09/edward-snowden-nsa-whistleblower-surveillance

Hall, K. (2016). Citizens don’t trust UK.GOV with their data. Retrieved October 29, 2016, from http://www.theregister.co.uk/2016/10/06/citizens_dont_trust_ukgov_with_their_data/

Harrison, D. A., Mykytyn, P. P., & Riemenschneider, C. K. (1997). Executive decisions about adoption of information technology in small business: Theory and empirical tests. *Information Systems Research*, *8*(2), 171–195. http://doi.org/10.1287/isre.8.2.171

Helmstaedter, M., Briggman, K. L., Turaga, S. C., Jain, V., Seung, H. S., & Denk, W. (2013). Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, *500*(7461), 168–174.

Hilbert, M., & Lopez, P. (2011). The world’s technological capacity to store, communicate, and compute information. *Science*, *332*(6025), 60–65. http://doi.org/10.1126/science.1200970

Horta, E. G., Castro, C. L. de, & Braga, A. P. (2015). Stream-based extreme learning machine approach for Big Data problems. *Mathematical Problems in Engineering*, *2015*, 1–17. http://doi.org/10.1155/2015/126452

IBM. (n.d.). Infographic: The four V’s of Big Data. Retrieved October 10, 2016, from http://www.ibmbigdatahub.com/infographic/four-vs-big-data

Introna, L., & Pouloudi, A. (1999). Privacy in the information age: Stakeholders, interests and values. *Journal of Business Ethics*, *22*(1), 27–38. http://doi.org/10.1023/A:1006151900807

Joyce, D. (2015). Privacy in the digital era: Human rights online. *Melbourne Journal of International Law*, *16*(1), 270–285.

Kaggle. (2014). *Higgs Boson Machine Learning challenge*. Retrieved from https://www.kaggle.com/c/higgs-boson

Kanter, A. S., Spencer, D. C., Steinberg, M. H., Soltysik, R., Yarnold, P. R., & Graham, N. M. (1999). Supplemental vitamin B and progression to AIDS and death in black South African patients infected with HIV. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *21*(3), 252–253.

Konrad, A. (2014). Airbnb cofounders to become first sharing economy billionaires as company nears $10 Billion valuation. Retrieved May 8, 2016, from http://www.forbes.com/sites/alexkonrad/2014/03/20/airbnb-cofounders-are-billionaires/#27e46b4d41ab

Korba, L., & Kenny, S. (2002). Towards meeting the privacy challenge: Adapting DRM. In *2002 ACM Workshop on Digital Rights Management, Held in Conjunction with the Ninth ACM Conference on Computer and Communications Security, November 18-22, 2002.* Washington, District of Columbia, USA. Retrieved from http://nparc.cisti-icist.nrc-cnrc.gc.ca/eng/view/object/?id=bcd2a953-61c6-46b4-b0ad-c01c2e00ee9b

Kuner, C. (2012). The European Commission's proposed data protection regulation: A copernican revolution in European Data Protection Law. *Bloomberg BNA Privacy and Security Law Report*, *6*(2012), 1–15.

Larson, R. G. (2013). Forgetting the First Amendment: How obscurity-based privacy and a right to be forgotten are incompatible with free speech. *Communication Law and Policy*, *18*(1), 91–120. http://doi.org/10.1080/10811680.2013.746140

Lashinsky, A. (2015). Uber: An oral history. Retrieved May 8, 2016, from http://fortune.com/2015/06/03/uber-an-oral-history/

Leung, M. K. K., Xiong, H. Y., Lee, L. J., & Frey, B. J. (2014). Deep learning of the tissue-regulated splicing code. *Bioinformatics*, *30*(12), i121–i129. http://doi.org/10.1093/bioinformatics/btu277

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry* (Vol. 75). New York, NY: Sage.

Ma, J., Sheridan, R. P., Liaw, A., Dahl, G. E., & Svetnik, V. (2015). Deep neural nets as a method for quantitative structure: Activity relationships. *Journal of Chemical Information and Modelling*, *55*(2), 263–274. http://doi.org/10.1021/ci500747n

Mangwanda, N. N. (2015). *The impact of the right to be forgotten on privacy and online information disclosure* (Unpublished master's thesis). Gordon Institute of Business Science, University of Pretoria, South Africa.

Mantelero, A. (2013). The EU Proposal for a general data protection regulation and the roots of the "right to be forgotten." *Computer Law & Security Review*, *29*(3), 229–235. http://doi.org/10.1016/j.clsr.2013.03.010

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big Data: The next frontier for innovation, competition and productivity*. Retrieved from http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

Martin, C. J. (2016). The sharing economy: A pathway to sustainability or a nightmarish form of neoliberal capitalism? *Ecological Economics*, *121*, 149–159. http://doi.org/10.1016/j.ecolecon.2015.11.027

McCracken, G. (1988). *The long Interview* (Vol. 13). New York, NY: Sage.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. New York, NY: Sage.

Miller, A. A. (2014). What do we worry about when we worry about price discrimination? The law and ethics of using personal information for pricing. *Journal of Technology Law & Policy*, *19*, 41.

Moore, G. E. (2006). Cramming more components onto integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp.114 ff. *IEEE Solid-State Circuits Newsletter*, *20*(3), 33–35. http://doi.org/10.1109/N-SSC.2006.4785860

Moses, L. B. (2007). Recurring Dilemmas: The law's race to keep up with technological change. *University of Illinois Journal of Law, Technology & Policy*, (2), 239–285. Retrieved from http://www.jltp.uiuc.edu/archives/moses.pdf

Narayanan, A., & Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy (sp 2008)* (pp. 111–125). IEEE. http://doi.org/10.1109/SP.2008.33

Newman, L. H. (2016). What we know about Friday's massive east coast Internet outage. Retrieved October 29, 2016, from https://www.wired.com/2016/10/internet-outage-ddos-dns-dyn/

Norman K. Denzin, Y. S. L. (2011). *The Sage handbook of qualitative research*. New York, NY: Sage.

O'Neil, C. (2016). Big Data algorithms are manipulating us all. Retrieved October 18, 2016, from https://www.wired.com/2016/10/big-data-algorithms-manipulating-us/

Ohm, P. (2010). Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review*, *57*, 1701.

Okori, W., & Obua, J. (2011). Machine Learning classification technique for famine prediction. In *Proceedings of the world congress on engineering* (Vol. 2, pp. 991–996).

Park, Y. (2016). 8 digital skills we must teach our children. Retrieved October 31, 2016, from https://www.weforum.org/agenda/2016/06/8-digital-skills-we-must-teach-our-children/

Patton, M. Q. (2005). Qualitative research. In *Encyclopedia of Statistics in Behavioral Science* (Vol. 3, pp. 344–347). Chichester, UK: John Wiley & Sons, Ltd. http://doi.org/10.1002/0470013192.bsa514

Peterson, M. J. (2008). *General Assembly*. *UN Doc A/RES/68/167* (Vol. 24). Oxford University Press. http://doi.org/10.1093/oxfordhb/9780199560103.003.0005

Pierce, E. L. (1857). *A treatise on american railroad law*. New York, NY: JS Voorhies.

Porter, M. E. (1991). Towards a dynamic theory of strategy. *Strategic Management Journal*, *12*(S2), 95–117. http://doi.org/10.1002/smj.4250121008

Premkumar, G. (2003). A meta-analysis of research on information technology implementation in small business. *Journal of Organizational Computing and Electronic Commerce*, *13*(2), 91–121. http://doi.org/10.1207/S15327744JOCE1302_2

Probst, C. W., Hunker, J., Gollmann, D., & Bishop, M. (2010). *Insider Threats in Cyber Security*. *Advances in Information Security* (Vol. 49). Boston, MA: Springer US. http://doi.org/10.1007/978-1-4419-7133-3

Protection of Personal Information Act (2013). Government Gazette. Retrieved from www.gov.za/documents/protection-personal-information-act

Rainie, L., & Duggin, M. (2016). *Privacy and information sharing*. *Pew Research Center Internet Project*. Retrieved from http://www.pewinternet.org/2016/01/14/privacy-and-information-sharing/

Ratti, C., Frenchman, D., Pulselli, R. M., & Williams, S. (2006). Mobile landscapes: Using location data from cell phones for urban analysis. *Environment and Planning B: Planning and Design*, *33*(5), 727–748. http://doi.org/10.1068/b32047

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., … Sabeti, P. C. (2011). Detecting novel associations in large data sets. *Science*, *334*(6062), 1518–1524. http://doi.org/10.1126/science.1205438

Robinson, N., Graux, H., Botterman, M., & Valeri, L. (2009). Review of EU data protection directive: Summary. *Information Commissioner's Office*.

Rogers, A. (2015). Google's search algorithm could steal the presidency. Retrieved May 7, 2016, from http://www.wired.com/2015/08/googles-search-algorithm-steal-presidency/

Rose, J., & Kalapesi, C. (2012). *Rethinking personal data: Strengthening trust*. *BCG Perspectives* (Vol. 16). Retrieved from http://www3.weforum.org/docs/WEF_IT_RethinkingPersonalData_Report_2012.pdf

Ross, P. K., & Blumenstein, M. (2015). Cloud computing as a facilitator of SME entrepreneurship. *Technology Analysis & Strategic Management*, *27*(1), 87–101. http://doi.org/10.1080/09537325.2014.951621

Rozovsky, L. E. (1974). Legal aspects of human and genetic engineering. *Man. LJ*, *6*, 291.

Rubinstein, I. S. (2013). Big Data: The end of privacy or a new beginning? *International Data Privacy Law*, *3*(2), 74–87. http://doi.org/10.1093/idpl/ips036

Rusk, N. (2015). Deep learning. *Nature Methods*, *13*(1), 35–35. http://doi.org/10.1038/nmeth.3707

Saunders, M., Lewis, P., & Thornhill, A. (2012). *Research methods for business students* (6th Ed). Harlow: Pearson Education Limited.

Saunders, M. N. K., & Lewis, P. (2014). *Doing research in business and management: An essential guide to planning your project*. New York, NY: Pearson Higher Education.

Schwab, K. (2016). *The Fourth Industrial Revolution: What it means and how to respond*. *World Economic Forum*. Retrieved from https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/

Schwandt, T. A. (1997). *Qualitative inquiry: A dictionary of terms*. New York, NY: Sage.

Schwartz, P. M. (2004). Property, privacy, and personal data. *Harvard Law Review*, *117*(7), 2056. http://doi.org/10.2307/4093335

Schwartz, P. M., & Solove, D. J. (2011). The PII problem: Privacy and a new concept of personally identifiable information. *New York University Law Review*, *86*, 1814.

Searls, D. (2013). *The Intention Economy: When customers take charge*. Boston, MA: Harvard Business Review Press.

Solove, D. J. (2007). ‘I’ve got nothing to hide’ and other misunderstandings of privacy. *San Diego Law Review*, *44*, 745.

Solove, D. J. (2006). A taxonomy of privacy. *University of Pennsylvania Law Review*, *154*(3), 477. http://doi.org/10.2307/40041279

South African Constitution (1996). Government Gazette. Retrieved from http://www.justice.gov.za/legislation/constitution/SAConstitution-web-eng.pdf

Spiekermann, S., Acquisti, A., Böhme, R., & Hui, K.-L. (2015). The challenges of personal data markets and privacy. *Electronic Markets*, *25*(2), 161–167. http://doi.org/10.1007/s12525-015-0191-0

Stewart, L. A. (2010). The impact of regulation on innovation in the United States: A cross-industry literature review. *Institute of Medicine*.

Stoycheff, E. (2016). Under surveillance: Examining Facebook's spiral o silence effects in the wake of NSA internet monitoring. *Journalism & Mass Communication Quarterly*, *93*(2), 296–311. http://doi.org/10.1177/1077699016630255

Sweeney, L. (2000). Simple demographics often identify people uniquely. *Health (San Francisco)*, *671*, 1–34.

Tatonetti, N. P., Denny, J. C., Murphy, S. N., Fernald, G. H., Krishnan, G., Castro, V., … Altman, R. B. (2011). Detecting Drug Interactions From Adverse-Event Reports: Interaction Between Paroxetine and Pravastatin Increases Blood Glucose Levels. *Clinical Pharmacology & Therapeutics*, *90*(1), 133–142. http://doi.org/10.1038/clpt.2011.83

Taylor, M. R., Rubin, E. S., & Hounshell, D. A. (2005). Control of SO^2 emissions from power plants: A case of induced technological innovation in the U.S. *Technological Forecasting and Social Change*, *72*(6), 697–718. http://doi.org/10.1016/j.techfore.2004.11.001

Taylor, R. C. (2010). An overview of the Hadoop/MapReduce/HBase framework and its current applications in bioinformatics. *BMC Bioinformatics*, *11*(Suppl 12), S1. http://doi.org/10.1186/1471-2105-11-S12-S1

Tene, O. (2011). Privacy: The new generations. *International Data Privacy Law*, *1*(1), 15–27. http://doi.org/10.1093/idpl/ipq003

Tene, O., & Polenetsky, J. (2012). To track or do not track: Advancing transparency and individual control in online behavioural advertising. *Minnesota Journal of Law, Science & Technology*, *13*, 281.

Tene, O., & Polonetsky, J. (2013). Big Data for all: Privacy and user control in the age of analytics. *Northwestern Journal of Technology and Intellectual Property*, *11*(5). Retrieved from http://scholarlycommons.law.northwestern.edu/njtip/vol11/iss5/1

The Huffington Post. (2009). Google's CEO on privacy: "If you have something you don't want anyone to know, maybe you shouldn't' be doing it'. Retrieved May 8, 2016, from http://www.huffingtonpost.com/2009/12/07/google-ceo-on-privacy-if_n_383105.html

Thong, J. Y. L., & Yap, C. S. (1995). CEO characteristics, organizational characteristics and information technology adoption in small businesses. *Omega*, *23*(4), 429–442. http://doi.org/10.1016/0305-0483(95)00017-I

Thornhill, J. (2016). Human rights challenged by desire for a crime-free world. Retrieved October 30, 2016, from https://www.ft.com/content/eb3d58f8-57c3-11e6-9f70-badea1b336d4

Tsatsou, P. (2011). Digital divides revisited: What is new about divides and their research? *Media, Culture & Society*, *33*(2), 317–331. http://doi.org/10.1177/0163443710393865

Underwood, R. H., & Cadle, R. G. (1996). Genetics, genetic testing, and the specter of discrimination: A discussion using hypothetical cases. *Kentucky Law Journal*, *85*, 665.

Venkatanathan, J., Karapanos, E., Kostakos, V., & Gonçalves, J. (2013). A network science approach to modelling and predicting empathy. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining - ASONAM '13* (pp. 1395–1400). New York, NY: ACM Press. http://doi.org/10.1145/2492517.2500295

Wesolowski, A., & Eagle, N. (2010). Parameterizing the dynamics of slums. In *AAAI Spring Symposium: Artificial Intelligence for Development*.

Westin, A. F. (1968). Privacy and freedom. *Washington and Lee Law Review*, *25*(1), 166.

Winn, J. K. (2016). *Consumer protection in the age of the "information economy."* New York, NY: Routledge.

World Economic Forum. (2011). *Personal Data: The Emergence of a new asset class. Forum American Bar Association*. Retrieved from http://www.weforum.org/reports/personal-data-emergence-new-asset-class

World Newsmedia Network. (2013). Big Data: The four V's. Retrieved October 30, 2016, from http://newsbizblog.blogspot.co.za/2013/11/big-data-four-vs.html

Xiong, H. Y., Alipanahi, B., Lee, L. J., Bretschneider, H., Merico, D., Yuen, R. K. C., … Frey, B. J. (2015). The human splicing code reveals new insights into the genetic determinants of disease. *Science*, *347*(6218), 1254806–1254806. article. http://doi.org/10.1126/science.1254806

Xiong, W., Yu, Z., Bei, Z., Zhao, J., Zhang, F., Zou, Y., … Xu, C. (2013). A characterization of Big Data benchmarks. In *2013 IEEE International Conference on Big Data* (pp. 118–125). IEEE. http://doi.org/10.1109/BigData.2013.6691707

Yin, R. K. (2009). *Case study research: Design and methods* (4th Ed.). Thousand Oaks, CA: Sage.

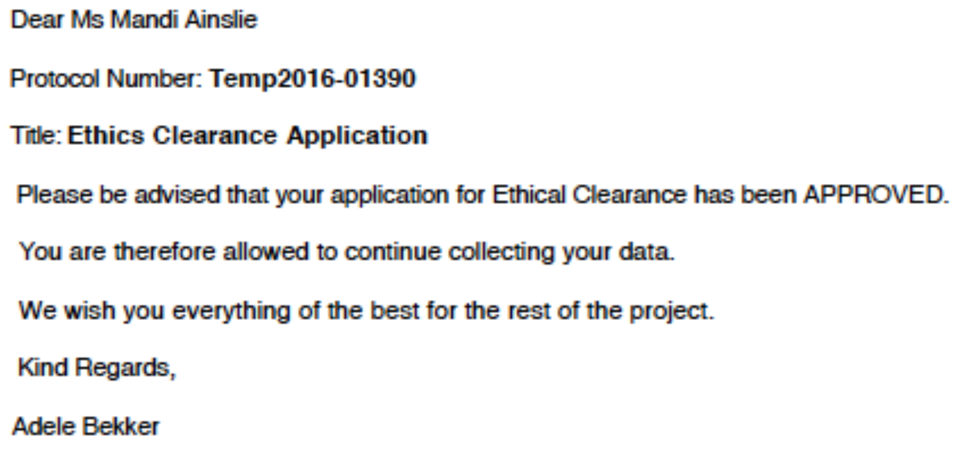
Yuan, E. J. (2013). A culturalist critique of "online community" in new media studies. *New Media & Society*, *15*(5), 665–679. http://doi.org/10.1177/1461444812462847

Zittrain, J. (2000). What the publisher can teach the patient: Intellectual property and privacy in an era of trusted privication. *Stanford Law Review*, *52*(5), 1201. http://doi.org/10.2307/1229513

Zyskind, G., Nathan, O., & Pentland, A. S. (2015). Decentralizing Privacy: Using Blockchain to Protect Personal Data. In *2015 IEEE Security and Privacy Workshops* (pp. 180–184). IEEE. http://doi.org/10.1109/SPW.2015.27

APPENDIX 1: ETHICAL CLEARANCE

A copy of the ethical clearance received for this research has been provided for reference purposes.



Dear Ms Mandi Ainslie

Protocol Number: **Temp2016-01390**

Title: **Ethics Clearance Application**

Please be advised that your application for Ethical Clearance has been APPROVED.

You are therefore allowed to continue collecting your data.

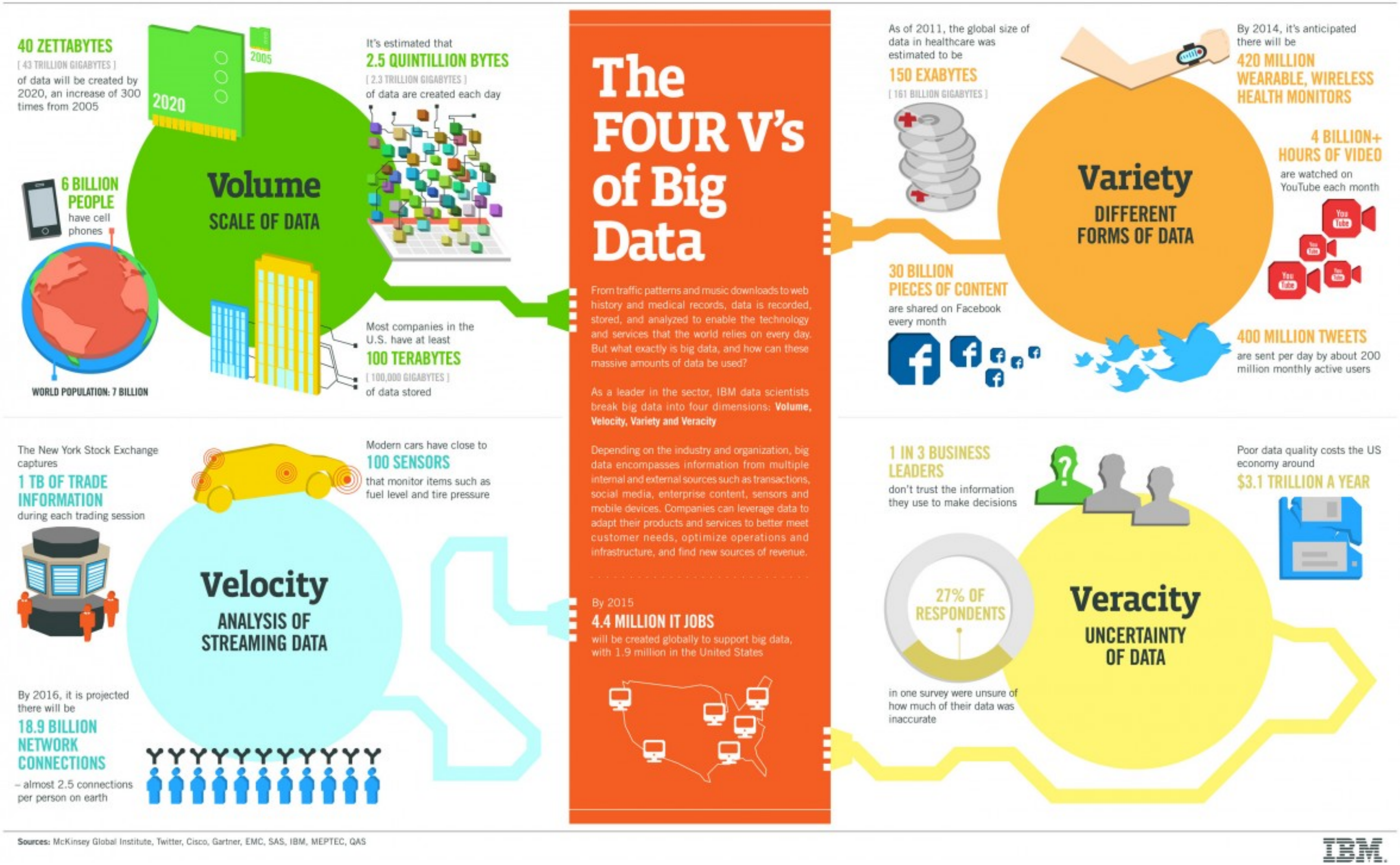
We wish you everything of the best for the rest of the project.

Kind Regards,

Adele Bekker

APPENDIX 2: THE FOUR V'S OF BIG `DATA

Retrieved from www.ibmbigdatahub.com/infographic/four-vs-big-data



The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015 **4.4 MILLION IT JOBS** will be created globally to support big data, with 1.9 million in the United States

Volume
SCALE OF DATA

40 ZETTABYTES [43 TRILLION GIGABYTES] of data will be created by 2020, an increase of 300 times from 2005

It's estimated that **2.5 QUINTILLION BYTES** [2.3 TRILLION GIGABYTES] of data are created each day

6 BILLION PEOPLE have cell phones

Most companies in the U.S. have at least **100 TERABYTES** [100,000 GIGABYTES] of data stored

Velocity
ANALYSIS OF STREAMING DATA

The New York Stock Exchange captures **1 TB OF TRADE INFORMATION** during each trading session

Modern cars have close to **100 SENSORS** that monitor items such as fuel level and tire pressure

By 2016, it is projected there will be **18.9 BILLION NETWORK CONNECTIONS** – almost 2.5 connections per person on earth

Variety
DIFFERENT FORMS OF DATA

As of 2011, the global size of data in healthcare was estimated to be **150 EXABYTES** [161 BILLION GIGABYTES]

By 2014, it's anticipated there will be **420 MILLION WEARABLE, WIRELESS HEALTH MONITORS**

4 BILLION+ HOURS OF VIDEO are watched on YouTube each month

30 BILLION PIECES OF CONTENT are shared on Facebook every month

400 MILLION TWEETS are sent per day by about 200 million monthly active users

Veracity
UNCERTAINTY OF DATA

1 IN 3 BUSINESS LEADERS don't trust the information they use to make decisions

Poor data quality costs the US economy around **$3.1 TRILLION A YEAR**

27% OF RESPONDENTS in one survey were unsure of how much of their data was inaccurate

Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC, QAS

APPENDIX 3: INTERVIEW SCHEDULE

QUALITATIVE INTERVIEW OUTLINE

Introduction:

Thank you for your time and providing your views and perception on the topic of the conflicting dynamism of Big Data strategy and privacy legislation.

Confirmation of confidentiality of data and reminder of participation is voluntary.
Explain the title of research and provide a brief overview.

Biographical and Background questions:

Briefly describe what the organisation does and the chosen industry it operates in.
Describe your role and responsibilities in the organisation.

Interview Question 1: Overview and Understanding:

1) Describe your involvement or understanding of Big Data and Business Intelligence.
   a) Availability of data
   b) High Speed Computation
   c) Computational Frameworks
2) Several Benefits of Big Data and Business Intelligence are noted, such as the impact on health care, geo location, Smart Grid technologies, traffic management retail, payment analysis and fraud analytics. What is your view on the benefits gained from Big Data.
3) How do you perceive your own privacy?
4) What are your thoughts on the statement that “if you have nothing to hide, you should therefore not care whether private companies or government agencies monitor and analyse online behaviour”?
5) What is your view on the criticism against Big Data and business intelligence?
   a) Incremental Effect
      - Netflix Recommendation Experiment, de-identified data was re-associated with identified individuals by cross referencing a de-identified database with publicly available resources accessible online. Once any piece of data has been linked to a person’s real identity, anonymity of the virtual identity is broken

b) Automated Decision Making / Behaviour Modification

- The influence of Machine Learning, predictive analytics that influences online behavioural advertising and personalisation applications, raises concerns around discrimination, self-determination and the narrowing of choice

c) Predictive Analytics

- Predictive analytics also has a useful application in law enforcement, national security, credit screening, insurance and employment.
  Pregnancy Example

d) Access and Exclusion

- Firstly, individuals exchange personal data for free services. Secondly, organisations are reluctant to share the valuable insights created by individual's personal data

6) What is your opinion on the efficacy (effective, useful etc.) of Privacy Legislation i.e. POPI?

7) Critics maintain that certain shortcomings are noted within all privacy legislation.

a) Firstly the reliance on Informed Choice of the individual. Individuals neither read nor understand privacy policies. What are your thoughts on the ambiguous language of these documents? Have you ever read the terms of use of Facebook or Google search engines?

b) The definition of Personal Identifiable Information (PII) is very broad and therefore, organisations use various methods of de-identification (anonymisation, pseudonymisation, encryption, key-coding, data sharing) to distance data from personal identities. What is you view on the validity of de-identification given that data scientist have proven that anonymised data can be re-identified.

c) The principal of data minimisation requires organizations to limit the collection of personal data to the minimum extent necessary to obtain their legitimate goals. However, Big Data incentivizes collection data for longer periods of time.

d) Lastly Privacy legislation has failed to keep pace with globalisation – the relentless improvement and expansion of tech capabilities and the changing ways in which individuals crate, share and use personal data. What is your view on globalisation of information?

e) You have any other critiques or concerns?

Interview question 2: Personal Data Store / Services

8) What is your view on moving data collection from the individual as a participant only to the **individual as the centre of personal data collection,** management

and use? It is organized around individuals collecting, storing, managing, using and sharing their own personal data for their own purposes.

a) A personal data store enabled via data warehousing technology and universal personal and mobile computing to help individuals manage personal data as a personal asset. This would include volunteered, observed and Inferred data

Interview question 3: Validity of Framework Pillars

The following 7 concepts would support a personal data store.

9) **Selective Disclosure.** The ability of customers to share their data selectively without disclosing more data than they wish to. Use Google example "My Activity" example and Personal Data Store example – change of address.

a) What do you think would the impact on selective disclosure be on the quality of data collected?

10) **Control over purpose and duration** – Building on selective disclosure, control over the purpose and duration of primary and secondary uses. This control may be achieved by "owner data agreement" and/or by technical means such as Data Rights Management (DRM) of meta-data tagging.

11) **Signalling** – the means for individuals to express demand for goods or services in open markets not tied to any single organisation. Use holiday / insurance example.

a) Would a vendor agree that this would enable them to better tailor their product / service propositions to a customer and promote higher revenue/sales?

12) **Identity management** – handle tasks such as the authentication and use of multiple identifiers while preventing correlation unless permitted by the user.

13) **Security**

a) What type of security considerations do you think this type of personal data store/service require?

14) **Data Portability** – ability to move all of one's data form one provider to another using standard data formats and interface protocols.

15) **Accountability and Enforcement** – accountability for protecting and securing personal data in accordance with the rights and permissions established by agreement and/or enforced by tagging mechanisms; and enforcement under self-regulatory guidelines and legal mandates, both back by comprehensive auditing.

Interview Question 4: Future Focused

16) Can you think of any other points that should be investigated?

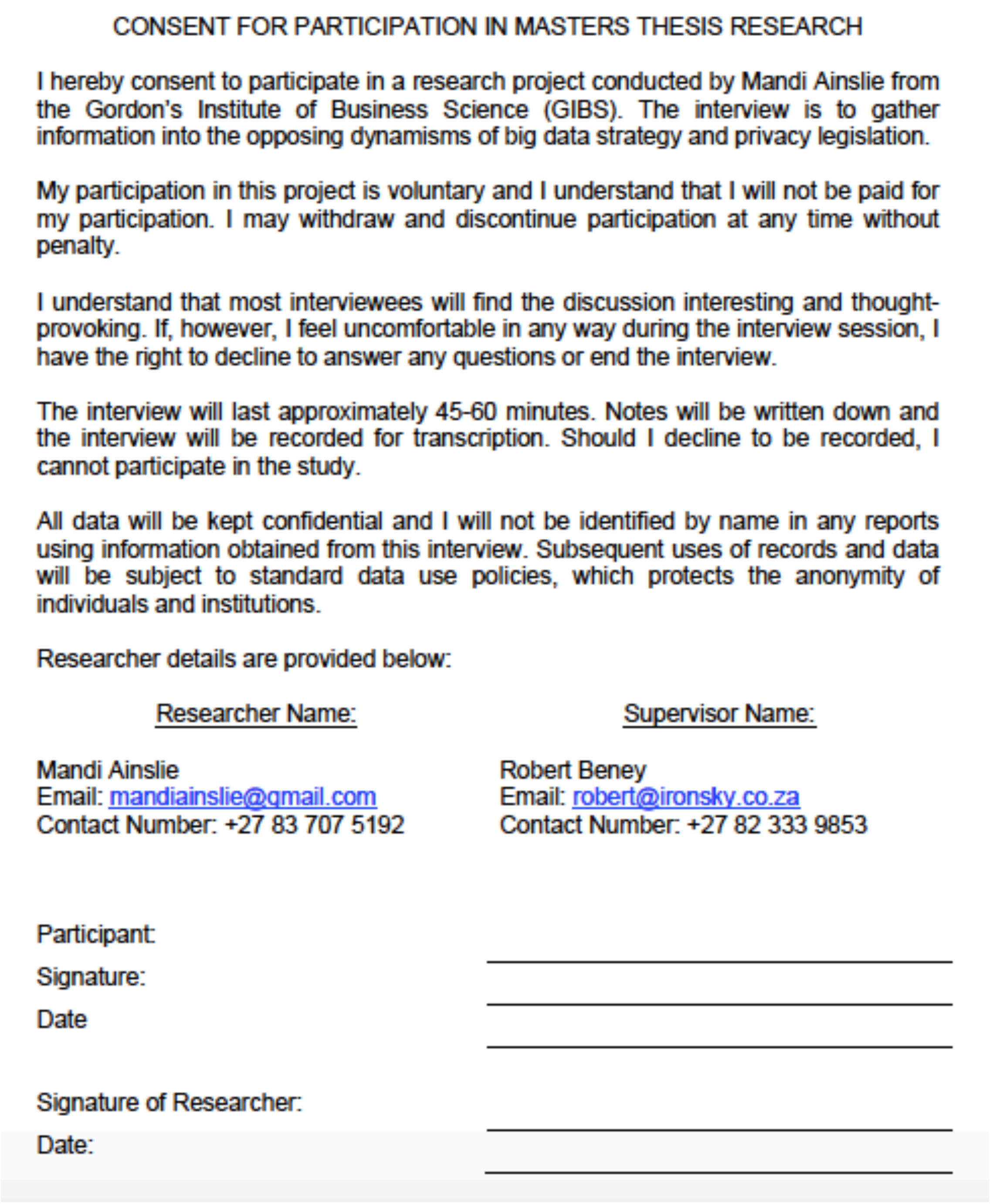
a) Technical feasibility – include data rights management and meta data tagging.

b) Intellectual coherence – Property Based theory – Selling / renting out your data

c) Existence of Business Incentives

Any Other comments on the points discussed?

APPENDIX 4: CONSENT FORM



CONSENT FOR PARTICIPATION IN MASTERS THESIS RESEARCH

I hereby consent to participate in a research project conducted by Mandi Ainslie from the Gordon's Institute of Business Science (GIBS). The interview is to gather information into the opposing dynamisms of big data strategy and privacy legislation.

My participation in this project is voluntary and I understand that I will not be paid for my participation. I may withdraw and discontinue participation at any time without penalty.

I understand that most interviewees will find the discussion interesting and thought-provoking. If, however, I feel uncomfortable in any way during the interview session, I have the right to decline to answer any questions or end the interview.

The interview will last approximately 45-60 minutes. Notes will be written down and the interview will be recorded for transcription. Should I decline to be recorded, I cannot participate in the study.

All data will be kept confidential and I will not be identified by name in any reports using information obtained from this interview. Subsequent uses of records and data will be subject to standard data use policies, which protects the anonymity of individuals and institutions.

Researcher details are provided below:

Researcher Name:	Supervisor Name:
Mandi Ainslie Email: mandiainslie@gmail.com Contact Number: +27 83 707 5192	Robert Beney Email: robert@ironsky.co.za Contact Number: +27 82 333 9853

Participant: ______________________________

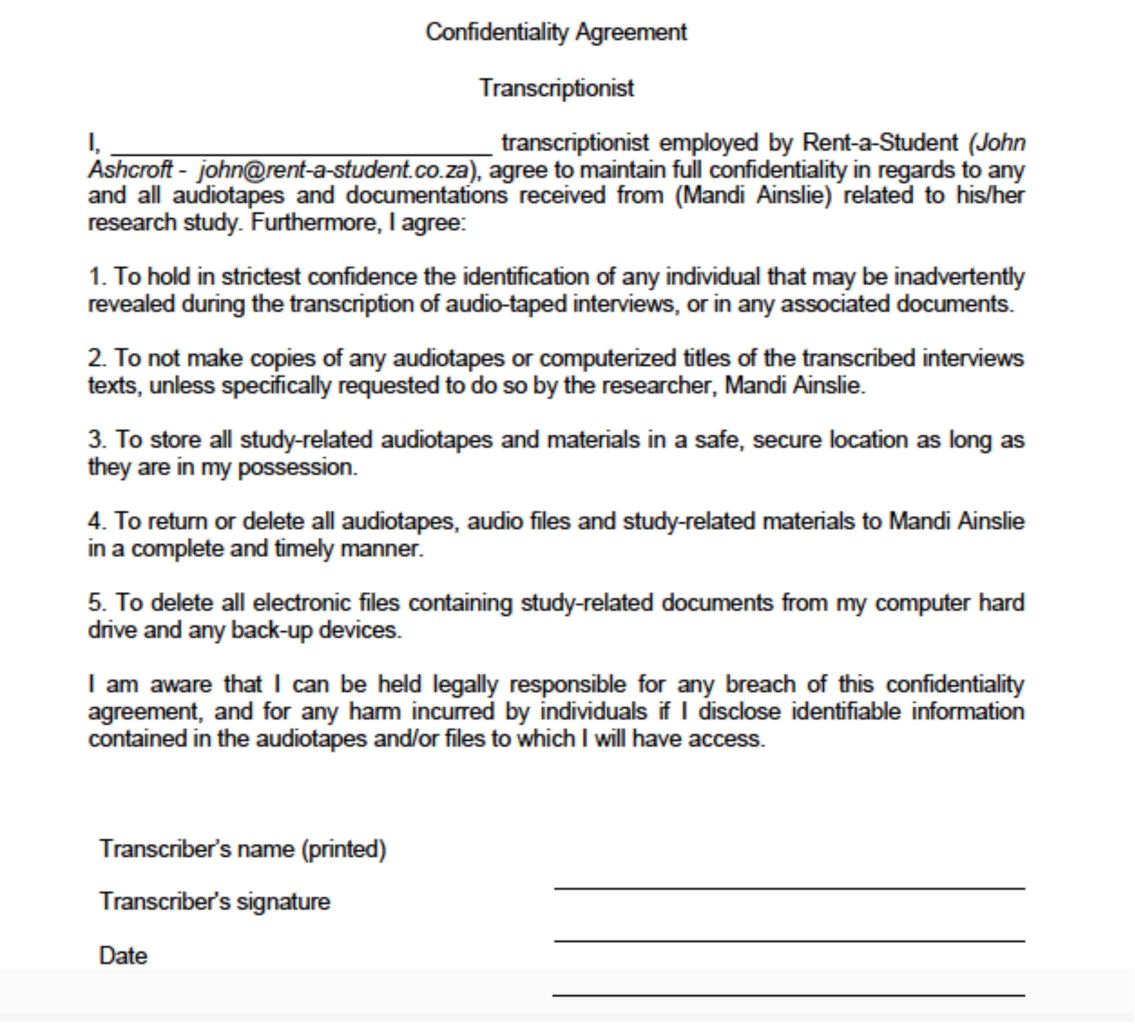
Signature: ______________________________

Date ______________________________

Signature of Researcher: ______________________________

Date: ______________________________

APPENDIX 5: CONFIDENTIALITY AGREEMENT



Confidentiality Agreement

Transcriptionist

I, ______________________________ transcriptionist employed by Rent-a-Student *(John Ashcroft - john@rent-a-student.co.za*), agree to maintain full confidentiality in regards to any and all audiotapes and documentations received from (Mandi Ainslie) related to his/her research study. Furthermore, I agree:

1. To hold in strictest confidence the identification of any individual that may be inadvertently revealed during the transcription of audio-taped interviews, or in any associated documents.

2. To not make copies of any audiotapes or computerized titles of the transcribed interviews texts, unless specifically requested to do so by the researcher, Mandi Ainslie.

3. To store all study-related audiotapes and materials in a safe, secure location as long as they are in my possession.

4. To return or delete all audiotapes, audio files and study-related materials to Mandi Ainslie in a complete and timely manner.

5. To delete all electronic files containing study-related documents from my computer hard drive and any back-up devices.

I am aware that I can be held legally responsible for any breach of this confidentiality agreement, and for any harm incurred by individuals if I disclose identifiable information contained in the audiotapes and/or files to which I will have access.

Transcriber's name (printed) ______________________________

Transcriber's signature ______________________________

Date ______________________________

APPENDIX 6: ATLAS.TI REPORT CODE GROUPS

Overall Understanding of Concepts

○ C01 - Overall understanding of Big Data

○ C01a - Big Data from a Privacy Lens*

Big Data Benefits

○ C02 - Benefits of Big Data

○ C02a - Value Proposition of Big Data*

○ C02ab - Understands a person's context*

○ C02b - Sustainable Innovation*

Criticisims Against Big Data

○ C05 - Criticism Against BD

○ C05a - Incremental Effect

○ C05b - Behaviour Modification

○ C05ba - Subliminal Messaging*

○ C05bb - Propaganda*

○ C05c - Predictive Analytics

○ C05d - Access and Exclusion

Personal Data Store Framework

○ C08 - Personal Data Store

○ C08b - Technical Feasibility*

○ C08ba - Dynamic Data*

○ C08bb - Static Data*

○ C09 - Selective Disclosure

○ C09a - Lower Operational Costs

○ C10 - Control over purpose and duration

○ C11 - Signalling

○ C11a - Higher Revenue for Company

○ C12 - Identity Management

○ C12a - Assumption of a digital footprint*

○ C13 - Security

○ C14 - Data portability

○ C15 - Accountability and Enforcement

○ C15a - Transparency*

Personal Privacy View

○ C03 - Personal Privacy

○ C04 - Nothing to hide

○ C04a - Human Right to privacy*

○ C04b - Ethics (Utilitarianism and Egoism)*

Privacy Legislation

○ C06 - Efficacy of Privacy Legislation

○ C06a - Law follows practice

○ C06ab - Work in conjunction with the law

○ C06b - Data Classification*

○ C07 - Shortcomings of Privacy Legislation

○ C07a - Informed Consent

○ C07b - Definition of PII

○ C07c - Data Minimisation

○ C07d - Globalisation

Additional Concepts

○ C16a - Intellectual Coherence (Property Based Theory)

○ C16b - Business Incentives

○ C17a - Intent*

○ C17b - Education*

○ C18 - Blockchain*