METHODOLOGY REPORT

Big Data Technology and National Security

Comparative International Perspectives on Strategy, Policy and Law

Law and Policy Program
Data to Decisions Cooperative Research Centre

June 2018

Research Team

Professor Louis de Koker, Program Leader
Professor Janet Chan, Project Leader
Professor Danuta Mendelson, Key Researcher
Associate Professor Lyria Bennett Moses, Key Researcher
Dr Alana Maurushat, Key Researcher
Mr David Vaile, Key Researcher
Mr Mike Gaffney, Researcher
Mr Gregory Sadler, Researcher
Mr Patrick Grierson, Researcher
Mr Daniel Cater, Project Research Assistant

Other research assistants
Ms Alana James
Ms Sonam Gordhan
Mr Jax Arnold

Interns (in alphabetical order)
Ms Kendy Ding
Mr Ciaran Finnane
Ms Monica Ma
Mr Kevin Tsu
Mr Atul Vidhata
Mr Vincent Wan
Ms Jacqueline Yip

Methodology Report

Authors
Chapter 1. Introduction – Janet Chan
Chapter 2. Background – Alana Maurushat, Lyria Bennett Moses
Chapter 3. Research Questions and Objectives – Janet Chan
Chapter 4. Research Methods and Sources of Data – Janet Chan, Lyria Bennett Moses
Chapter 5. Indicators of a Legal And Policy Framework that Supports 'Desirable And Effective' Big Data Practices – Louis de Koker, David Vaile, Lyria Bennett Moses, Alana Maurushat, Danuta Mendelson

Design of questionnaires and sample
Janet Chan, Lyria Bennett Moses, Alana Maurushat, Louis de Koker

Technical editing
Mr David Vaile

Other Reports from this Project

Australia Report
UK Report
Canada Report
Comparative Report
Select Bibliography
Technical References [for Australia Report]

Contents

List of Tables

List of Abbreviations and Acronyms

These are mainly Australian unless otherwise noted. Legislation is Australian Commonwealth jurisdiction unless noted.

AAA	Australian Airports Association
AAT	Administrative Appeals Tribunal
ABF	Australian Border Force
ACAT	Aviation Criminal Assessment Team
ACBPS	Australian Customs and Border Protection Service
ACC	Australian Crime Commission
ACC Act	*Australian Crime Commission Act 2002*
ACC Board	Australian Crime Commission Board
ACID	Australian Criminal Intelligence Database
ACIF	Australian Criminal Intelligence Forum
ACIM	Australian Criminal Intelligence Model
ACIMS	Australian Criminal Intelligence Management Strategy
ACLEI	Australian Commission for Law Enforcement Integrity
ACPO	Association of Chief Police Officers (UK)
ACTC	Australian Counter Terrorism Centre
ADF	Australian Defence Force
ADR	alternative dispute resolution
AEMVF	Australian Emergency Management Volunteer Forum
AFAC	Australasian Fire and Emergency Service Authorities Council
AFP	Australian Federal Police
AFPA	Australian Federal Police Association
AGD	Attorney-General's Department
AGDRP	Australian Government Disaster Recovery Payment
AHRC	Australian Human Rights Commission
AIC	Australian Intelligence Community
AIPJ	Australia Indonesia Partnership for Justice
ALEIN	Australian Law Enforcement Intelligence Network
All-in	'All-in' airport policing model
ALRC	Australian Law Reform Commission
AMTA	Australian Mobile Telecommunications Association
ANAO	Australian National Audit Office
Anderson Report	A Question of Trust: Report of the Investigatory Powers Review, 2015 (UK)
ANPR	Automatic Number Plate Recognition
ANU	Australian National University
ANZPAA	Australian New Zealand Policing Advisory Agency
ANZTC	Australia-New Zealand Counter-Terrorism Committee
APCERT	Asia Pacific CERT (computer emergency response team)
APEC	Asia-Pacific Economic Cooperation
APP	Australian Privacy Principles
APS	Australian Public Service
APSC	Australian Public Service Commission
AQIS	Australian Quarantine Inspection Service
ASH	All Source Hub, Metropolitan Police (UK)
ASIC	Australian Securities and Investments Commission
ASIO	Australian Security Intelligence Organisation
ASIS	Australian Secret Intelligence Service
ASU	Australian Services Union
ATA	*Anti-Terrorism Act 2015* (Canada); also, Bill C-51
ATO	Australian Taxation Office
ATSA	*Aviation Transport Security Act 2004*
ATSILS	Aboriginal and Torres Strait Islander Legal Service
AusAID	Australian Agency for International Development
AusCheck	Australian Background Checking Service
AUSTRAC	Australian Transaction Reports and Analysis Centre
AvSec	Aviation Security Service (New Zealand)
BCR	Building Community Resilience Grants
Beale Audit	Federal Audit of Police Capabilities
BPD	Bulk Personal Datasets (UK)
C-51, Bill C-51	See *Anti-Terrorism Act 2015* (Canada)
CAC	Communications Access Co-ordinator
CAC Act	*Commonwealth Authorities and Companies Act 1997*
CAF	Canadian Armed Forces
CAP-AU-STD	Common Alerting Protocol - Australian Profile
CASA	Civil Aviation Safety Authority
CBSA	Canada Border Services Agency
CCC	Crisis Coordination Centre
CCS	Children's Contact Service
CCTV	Closed-circuit television
CDPP	Commonwealth Director of Public Prosecutions

CEF	Container Examination Facility
CEG	Communications Exploitation Group, Metropolitan Police (UK)
CEPO	Community Engagement Police Officer
CERT	Computer Emergency Response Team
CESG	Communications-Electronics Security Group, a part of GCHQ (UK)
CFIA	Canada Food Inspection Agency
CIPMA	Critical Infrastructure Program for Modelling and Analysis
CIR	Critical Infrastructure Resilience
CJEU	Court of Justice of the European Union
CLCS	UN Commission on the Limits of the Continental Shelf
CNI	Certificates of No Impediment
CNS	Canadian Nuclear Safety
COAG	Council of Australian Governments
COBRA	Classification Operations Branch Records Administration
Commonwealth Framework	Commonwealth Organised Crime Strategic Framework
COMSEC	communications security (Canada)
CPGs	Commonwealth Procurement Guidelines
CRA	Canada Revenue Agency
CRC	Cooperative Research Centre
CRCC	Civilian Review and Complaints Commission for the RCMP
CRIMELINK	PNC application (UK)
CrimTrac	CrimTrac
CSE	Communications Security Establishment [Canada]
CSIS	Canadian Security Intelligence Service
CSP	Carriage Service Provider
CTIRU	Counter Terrorism Internal Referral Unit, Metropolitan Police (UK)
Customs Act	*Customs Act 1901*
Customs	Australian Customs and Border Protection Service (now DIBP/ABF)
CVE	countering violent extremism
CVESC	Countering Violent Extremism Sub-Committee
DBCDE	Department of Broadband, Communications and the Digital Economy
DFAIT – T&D	Department of Foreign Affairs, Trade and Development (Canada)
DFAT	Department of Foreign Affairs and Trade
DHS	Department of Homeland Security (US)
DIBP	Dept of Immigration and Border Protection
Digital Rights Ireland case	CJEU, Joined Cases C-293/12 and C-594/12 *Digital Rights Ireland* and *Seitlinger and others*, EU:C:2014:238
DNA	deoxyribonucleic acid (genetic material)
DOD	Department of National Defence (Canada)
DOORS	Detection Of Overall Risk Screen
DRIPA 2014	Data Retention and Investigatory Powers Act 2014 (UK)
DVS	Document Verification Service
ECHR	European Court of Human Rights
ECJ	European Court of Justice
EM	Explanatory Memorandum
EPIC	Electronic Privacy Information Center (US NGO)
EU	European Union
EWB	economic well-being, one of GCHQ's purposes (UK)
FaHCSIA	Department of Families, Housing, Community Services and Indigenous Affairs
FASO	family and sexual offences
Five Eyes	UK, USA, Canada, Australia and New Zealand intelligence communities
FMA Act	*Financial Management and Accountability Act 1997*
FOI	Freedom of Information
FTRACC	Financial Transactions and Reports Analysis Centre of Canada
Fusion	Criminal Intelligence Fusion Capability
GCHQ	Government Communications Headquarters (UK)
Harper	former Canadian PM 2006–2015 Stephen Harper
HMIC	Her Majesty's Inspectorate of Constabulary (UK)
HOCOLEA	Heads of Commonwealth Operational Law Enforcement Agencies
HRA	*Human Rights Act 1988* (UK)
IBAC	Independent broad-based anti-corruption commission
IC	Information Commissioner (UK)
ICAC	South Australian Independent Commissioner Against Corruption
ICAO	International Civil Aviation Organisation
ICC	International Criminal Court
ICCPR	International Covenant on Civil and Political Rights
ICL	Independent Children's Lawyer
ICS	Integrated Cargo System
ICO	Information Commissioner's Office (UK)
IDS	intrusion detection systems

IPS	intrusion prevention systems
IGIS	Inspector General of Intelligence and Security
ILSAC	International Legal Services Advisory Council
ILUA	Indigenous Land Use Agreement
IMO	International Maritime Organisation
IMSI	International Mobile Subscriber Identity (subscriber or device identifier stored in a SIM card)
InSeC	Intelligence Services Commissioner (UK)
INTEL	intelligence gathering
IOCCO	Interception of Communications Commissioner (UK)
IP	Internet Protocol, hence IP address
IPA	Institute of Public Affairs
IPB	Investigatory Powers Bill (UK)
IPC	Investigatory Powers Commissioner, proposed in IPB (UK)
IPT	Investigatory Powers Tribunal (UK)
ISC	Intelligence and Security Committee of Parliament (UK)
ISC Report	*Privacy and Security: A modern and transparent legal framework,* March 2015 (UK)
ISCom	Intelligence Services Commissioner (UK)
ISIC	Independent Surveillance and Intelligence Commission, new body proposed by Anderson Report (UK)
ISOC-AU	Internet Society of Australia (now Internet Australia, IA)
ISP	Internet Service Provider (a Carriage Service Provider or CS Intermediary)
ISPS Code	International Ship and Port Facility Security Code
ISS	International Social Service Australia
ITAC	Integrated Terrorism Assessment Centre (Canada)
ITSA	Insolvency and Trustee Service Australia
JAIG	Joint Aviation Intelligence Group
JAIT	Joint Aviation Investigation Team
JIC	Joint Intelligence Committee (UK)
KCLS	Kimberley Community Legal Service
Law Council	Law Council of Australia
LBS	location-based solution
LE	law enforcement
LSMUL	Legal Services Multi-Use List
Maritime regulations	Maritime Transport and Offshore Facilities Regulations 2003
MCPEMP	Ministerial Council for Police and Emergency Management—Police
MEAA	Media Entertainment and Arts Alliance
MI5	Security Service (UK)
MI6 (SIS)	Secret Intelligence Services (UK)
MITS	Operational Security Standard: Management of Information Technology Security (Canada)
MoU	Memorandum of Understanding
MSIC	Maritime Security Identification Card
MTOFSA	*Maritime Transport and Offshore Facilities Security Act 2003*
MUA	Maritime Union of Australia
NADRAC	National Alternative Dispute Resolution Advisory Council
NBO	National Back Office of NHS (UK)
NCA	National Crime Agency (UK)
NCCCP	National Crisis Coordination Capability Program
NCS	National Classification Scheme
NCTC	National Counter-Terrorism Committee
NCTR	National Criminal Target Report
NDA	National Defence Act [Canada]
NDRP	Natural Disaster Resilience Program
NDRRA	Natural Disaster Relief and Recovery Arrangements
NEMC	National Emergency Management Committee
NEMVAP	National Emergency Management Volunteer Action Plan
NGO	Non-government organisation
NHS	National Health Service (UK)
NIINA	National Information and Intelligence Needs Analysis
NISS	National Identity Security Strategy
NNTT	National Native Title Tribunal
NOCRP	National Organised Crime Response Plan
NPRS	National Police Reference System
NS	national security, one of GCHQ's purposes (UK)
NSA	National Security Agency (US)
NSC	National Security Council (UK)
NSCDD	National Security Capability Development Division
NSDR	National Strategy for Disaster Resilience
NSW	New South Wales
NT	Northern Territory
NTS	National Target System
O'Connor	*O'Connor Recommendations* (Canada)
OAIC	Office of the Australian Information Commission
OCA	*Organised Crime in Australia* Report (various years)
OCRP	Organised Crime Response Plan
OCSEC	Office of the Communications Security Establishment Commissioner (Canada)

OCSF	Organised Crime Strategic Framework
OCTA	Organised Crime Threat Assessment
OLDP	Office of Legislative Drafting and Publishing
OLSC	Office of Legal Services Coordination
ONA	Office of National Assessments
OPC	Office of Parliamentary Counsel
OPCAT	Optional Protocol to the Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment
OPP	Office of the Public Prosecutor in Papua New Guinea
OSC	Office of Surveillance Commissioners (UK)
OTS	Office of Transport Security
PAES	Portfolio Additional Estimates Statements
Patriot Act	*USA Patriot Act of 2001* (Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act)
PBS	Portfolio Budget Statements
PDRR	prevention, detection, response and recovery (Canada)
PERC	Royal Canadian Mounted Police External Review Committee
PFA	Police Federation of Australia
PHBR	Parliament House Briefing Room
PIA	Privacy Impact Assessment
PI	Privacy International
PILON	Pacific Island Officers' Network
PIM	Victorian Public Interest Monitor
PIPEDA	*Personal Information Protection and Electronic Documents Act 2000* (Canada)
PJC-ACC	Parliamentary Joint Committee on the Australian Crime Commission
PJCIS	Parliamentary Joint Committee on Intelligence and Security
PJCLE	Parliamentary Joint Committee on Law Enforcement
PNC	Police National Computer (UK)
PND	Police National Database (UK)
PNG	Papua New Guinea
Poole ruling	*Jenny Paton and others vs. Poole Borough Council*, IPT/09/01 (UK)
PPATK	Indonesian Government's Financial Intelligence Unit
PPS	Personal Property Securities
PPSA	*Personal Property Securities Act 2009*
PPSR	Personal Property Securities Register
PSEPC	Public Safety and Emergency Preparedness Canada
PSM	Public Service Medal
PSO	AFP Protective Services Officer
QUEST	Querying Using Enhanced or Extended Search Techniques, PNC application (UK)
RAP	Reconciliation Action Plan
RCMP	Royal Canadian Mounted Police
RFI	Request for Information
RIPA	*Regulation of Investigatory Powers Act 2000* (UK)
RIS	Regulation Impact Statement
RUSI	Royal United Services Institute (UK)
RUSI Report	*A Democratic Licence to Operate*, July 2015 (UK)
SA	South Australia
SACL	Sydney Airport Corporation Limited
SACS	social and community services
SC	prevention and detection of serious crime, one of GCHQ's purposes (UK)
SCAG	Standing Committee of Attorneys-General
SCISA	*Security of Canada Information Sharing Act* 2015
SCLJ	Standing Council on Law and Justice
SCPEM	Standing Council on Police and Emergency Management
SCPEM	Standing Council on Police and Emergency Management
SCC	Surveillance Camera Commissioner (UK)
SES	Senior Executive Service
SIAs	surveillance and intelligence agencies
SIEM	security incident and event management
SIRC	Security Independent Review Committee (Canada)
SIRC Act	*Security Independent Review Committee Act* (Canada)
SIS	Secret Intelligence Services, a.k.a. , MI6 (UK)
SLCAC	Senate Legal and Constitutional Affairs Committee
SLF Program	Strengthening Legal Frameworks to Counter Terrorism Program
Smith Review	*Review of Homeland and Border Security*
SOCN	Serious and organised crime networks
SOG on OC	Senior Officers' Group on Organised Crime
TBC	Treasury Board of Canada
TIA Act	*Telecommunications (Interception and Access) Act 1979*
TISN	Trusted Information Sharing Network
TOR	The Onion Router (encrypted public network)
TRAM	Threat Risk Assessment Methodology
TWU	Transport Workers Union
UN	United Nations

UNCAC	United Nations Convention against Corruption
UNCITRAL	United Nations Commission on International Trade Law
UNODC	United Nations Office on Drugs and Crime
UPM	Universal Policing Model
URL	Uniform resource locator [translated from IP address by DNS]
VI	Victorian Inspectorate
VIC	Visitor Identification Card
VODS	Vehicle Online Descriptive Search, PNC application (UK)
VOIP	Voice over Internet Protocol
VPN	Virtual Private Network
WA	Western Australia
WHS	work[place] health and safety
WIPO	World Intellectual Property Organization
WPSS	Wireless Priority Service System
WTO	World Trade Organization

1. INTRODUCTION

1.1 The Research Project

This report provides an overview of the questions and methodology of the research project 'Big Data Technology and National Security: Comparative International Perspectives on Strategy, Policy and Law in Australia, the United Kingdom and Canada' conducted by the Law and Policy Research Program of the Data to Decisions CRC.

Research findings on the three countries are reported separately in three country reports (Australia Report, UK Report and Canada Report). Conclusions and recommendations arising from research on the three jurisdictions are contained in the Comparative Report.

The main aim of the project is to examine the policies, regulatory approaches, processes and strategies used by these countries to balance the management and exploitation of Big Data for law enforcement and national security purposes, while safeguarding confidentiality and security of sensitive personal information, as well as the accuracy of data sets. The focus of the project is on members of the Five Eyes intelligence community.

This set of reports stem from an independent study funded by the Data to Decisions Cooperative Research Centre and undertaken by a team of law and policy researchers from Deakin Law School and UNSW Law under the lead of Professor Janet Chan (UNSW). Officials of the Attorney-General's Department joined the research team to provide logistical support.

The study was undertaken for the Attorney-General's Department but was conducted independently. The involvement of government officials in the logistics of the study, the empirical interviews or the review process should not be interpreted as an expression of any view on the study or its findings, either by government agencies or any of their officials.

1.2 Organisation of the Report

The rest of the report is organised as follows.

Chapter 2 provides a review of the existing research literature, summarising the key issues regarding the use of Big Data in general, and its use in the context of national and international security in particular.

Chapter 3 details the key research questions that are addressed in the case studies and summarises the overall research aims of the project.

Chapter 4 outlines the research methods, sources of data and the research process involved in conducting the project.

Chapter 5 introduces a framework that the researchers have established for analysing the governance and regulation of the use of Big Data for national security.

2. BACKGROUND

The capture and analysis of data is experiencing exponential growth. Advancements in data capture and storage technologies mean that more data is available to use and analyse. Datasets that may be available to government include but are not limited to open government datasets (‘**Open Data’**), public datasets such as social media data, metadata, and datasets in new fields such as the ‘Internet of Things,’ or sensors in Smart Cities.[1] ‘Big Data’ is a term that is used to describe large and often disparate datasets that are able to be analysed using faster processing and smarter (autonomous and semi-autonomous) analytics.

The idea of Big Data has emerged from this growth and the proliferation of digitised information. Innovative technologies and processes are increasing the ability to accurately and speedily assess, interpret and understand Big Data in a variety of fields, professions and contexts. In a modern setting, Big Data and its analysis can play crucial roles in both commercial and government contexts; for example in the delivery of government services such as healthcare and transportation.

For the purposes of this project it is important to note that Big Data is also impacting the ways in which data is analysed and used to predict, investigate, understand and disrupt crime and other incidents in the fields of law enforcement and intelligence.

While the project is concerned primarily with the use of Big Data for law enforcement, defence and intelligence in the context of national security, the use of Big Data analytics for other more general purposes can provide useful insights about its implications. This chapter will, therefore, refer to general principles and implications of Big Data, as well as to specific legal and policy issues arising when Big Data is used by law enforcement and intelligence in the context of ‘national security’.

2.1 Defining Big Data

This chapter will distinguish between three aspects of Big Data. It will refer to vast volumes of actual data as ‘Big Data’, Big Data software and technologies as ‘Big Data tools’, and the analysis of Big Data simply as ‘data analytics.’ Big Data tools are touted for bringing smarter analytics that provide richer and potentially more useful insights.

‘Big Data’ is a term used to describe new phenomena of data and data relationships being discovered by continuously evolving technology. It is believed that the term ‘Big Data’ was first coined in 1997 by NASA scientists Michael Cox and David Ellsworth in an attempt to describe the problem with visualisation of data within computer systems:

> data sets are generally quite large, taxing the capacities of main memory, local disk, and even remote disk. We call this the problem of Big Data. When data sets do not fit in main memory (in core), or when they do not fit even on local disk, the most common solution is to acquire more resources.[2]

By 1998 John Masey, Chief Data Scientist at SGI, gave a paper entitled ‘Big Data… and the Next Wave of Infrastress’[3] at a USENIX meeting. This paper articulated some of the essential components of Big Data – volume of data; collection and storage; and the need for better

[1] R Kitchin, ‘The Real-Time City? Big Data and Smart Urbanism’ (2013) 79 *Geo-Journal* 1; M Batty, ‘Big Data, Smart Cities and City Planning,’ (2013) 3(3) *Dialogues in Human Geography*.
[2] M Cox and D Ellsworth, ‘Application-controlled Demand Paging for Out-of-Core Visualization’, *Proceedings of the 8th Conference on Visualization* (1997), IEEE Computer Society Press, 235–244.
[3] See <http://static.usenix.org/event/usenix99/invited_talks/mashey.pdf>.

data analytic tools.[4] Since then, Big Data has been described as a data revolution.[5] However, as one expert put it, 'the real revolution is not in the machines that calculate data but in data itself and how we use it.'[6] Given its revolutionary impacts, Big Data is often labelled as a 'disruptive innovation'.[7] Like other disruptive innovations, it can re-shape the landscape of benefits and risks arising from its use, and their distribution among affected parties.

Big Data relates to other more historical terms such as 'informatics', 'data mining', 'smart analytics', 'data science', 'data analytics', 'predictive analytics,' 'artificial intelligence', 'machine-learning', 'open data' and 'cloud computing' to name but a few. Big Data, however, represents something more than this.

Big Data trends, especially in the areas of governmental use, are often linked to Open Data initiatives.[8] Governments and organisations are increasingly opening up their datasets, making data available and accessible, reusable and redistributable, typically for any purpose. Many Big Data tools use open or publicly available data and link to other datasets which may be open or private.[9] Big Data and Open Data are therefore parallel and mutually-reinforcing trends.

Proposed definitions of Big Data from industry, governments (US, EU and Australia) and organisations are listed in Table 2-1.

[4] G Press, 'A Very Short History of Big Data' (May 9, 2013) *Forbes* <http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/>.
[5] Victor Mayer-Schönberger and K Cukier, *Big Data: A Revolution that will Transform How We Live, Work, and Think* (John Murray Publishers, 2013).
[6] Victor Mayer-Schönberger and K Cukier, *Big Data: A Revolution that will Transform How We Live, Work, and Think* (John Murray Publishers, 2013), 7.
[7] N Cortez, 'Regulating Disruptive Innovation' (2014) 20 *Berkeley Technology Law Journal*.
[8] R Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences* (Safe, London, 2014); Open Data Initiative, *Global Open Data Initiative* (June, 2013) <http://globalopendatainitiative.org/>; David Vaile, Alana James, Lyria Bennett-Moses, Louis De Koker and Alana Maurushat, *Review of Barriers to Open Data and Related Re-use of Information in Five Exemplar Federal Data Sets* (November 2014), report available on file with authors.
[9] Open Data Handbook, *What is Open Data* <http://opendatahandbook.org/guide/en/>.

Table 2-1 Assorted Definitions of Big Data

Source	Definitions
Executive Office of the President of the United States, White Paper, 'Big Data: Seizing Opportunities, Preserving Values' May 2014.	There are many definitions of 'Big Data' which may differ depending on whether you are a computer scientist, a financial analyst, or an entrepreneur pitching an idea to a venture capitalist. Most definitions reflect the growing technological ability to capture, aggregate, and process an ever-greater volume, velocity, and variety of data. In other words, 'data is now available faster, has greater coverage and scope, and includes new types of observations and measurements that previously were not available.'[10] More precisely, Big Datasets are 'large, diverse, complex, longitudinal, and/or distributed datasets generated from instruments, sensors, Internet transactions, email, video, click stream, and/or all other digital sources available today and in the future.'[11]
European Big Data Value Strategic Research and Innovation Agenda, July 2014.	Economic and social activities have long relied on data. But today the increased volume, velocity, variety, and social and economic value of data signals a paradigm shift towards a data-driven socioeconomic model. In parallel with the continuous and significant growth of data has come better data access, availability of powerful ICT systems, and ubiquitous connectivity of both systems and people. This has led to intensified activities around Big Data and Big Data Value. Powerful tools have been developed to collect, store, analyse, process, and visualize huge amounts of data. Open data initiatives have been launched to provide broad access to data from the public sector, business and science.
SAS and e-Skills UK, 'Big Data Analytics: An Assessment of Demand for Labour and Skills 2012–2017,' January 7, 2013.	There is currently no singular, internationally recognised definition of what constitutes 'Big Data'. Many reports make reference to the three 'V's proposed in 2001 by the META Group, i.e. Volume (a reference to data stores of petabytes or above), Velocity (the requirement for real-time collection/analysis of data) and Variety (generation of data in diverse formats from a variety of collection mechanisms), and, in some cases, this definition has been further expanded to incorporate related considerations such as Variability (temporal data peaks) and Complexity (issues relating to linking/cleaning/editing data from different sources) for example. In all cases, however, the terminology employed to describe Big Data is not an operational one and, as such, cannot be used to identify a distinct sector, occupation, process, etc. In fact, even the core terms are highly subjective and liable to change in accordance with social/technological developments.
David Vesset, 'Worldwide Big Data Technology and Services 2012-2015 Forecast,' IDC Analysis for the European Union (March 2012).	IDC defines Big Data technologies as a new generation of technologies and architectures designed to extract value economically from very large volumes of a wide variety of data by enabling high-velocity capture, discover, and/or analysis.
Boyd and Crawford, Council for Big Data Ethics and Society, 2014.[12]	*Technology*: maximizing computation power and algorithmic accuracy to gather, analyze, link, and compare large data sets. *Analysis*: drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims. *Mythology*: the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy.

Source	Definitions
Peter Mell, 'NIST Presentation: Overview of Big Data and Security Implications' National Institute of Standards and Technology, 2015	Big Data is where the data volume, acquisition velocity, or data representation limits the ability to perform effective analysis using traditional relational approaches or requires the use of significant horizontal scaling for efficient processing.
Australian Public Service Better Practice Guide for Big Data 2015	The data analysis being undertaken uses a high volume of data from a variety of sources including structured, semi-structured, unstructured or even incomplete data; and The size (volume) of the data sets within the data analysis and velocity with which they need to be analysed has outpaced the current abilities of standard business intelligence tools and methods of analysis.
The Australian Public Service Big Data Strategy 2013	Big Data refers to the vast amount of data that is now being generated and captured in a variety of formats and from a number of disparate sources.

While there is no universally agreed rigorous definition of Big Data within the industry, the term generally denotes the following characteristics: volume, source variety, processing capacity, and smart analytics (see Table 2-2).

[10] Liran Einav and Jonathan Levin, 'The Data Revolution and Economic Analysis,' Working Paper, No. 19035, National Bureau of Economic Research (2013) <http://www.nber.org/papers/w19035>, in Executive Office of the President of the United States, White Paper, *Big Data: Seizing Opportunities, Preserving Values* (May 2014), 2–3.

[11] National Science Foundation, 'Solicitation 12–499: Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA)', 2012 <http://www.nsf.gov./pubs/2012/snf12499.pdf> in Executive Office of the President of the United States, White Paper, *Big Data: Seizing Opportunities, Preserving Values*, (May 2014), 2–3.

[12] d boyd and K Crawford, 'Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon' (2012) 15 *Information, Communication & Society* 662, 663; Council for Big Data Ethics and Society (2014) <http://bdes.datasociety.net/>.

Table 2-2 Characteristics of Big Data

VOLUME	Vast amounts or volume of information, often from disparate and multiple databases, capable of collection. The volume is so large that there isn't a mathematical term for this other than to say that it is a Google +;
SOURCE VARIETY	Uses a wide range of sources and formats of data where datasets grow in volume, variety and complexity. Big Data increasingly draws on open datasets.
PROCESSING CAPACITY	Is able to process a variety of datasets with large volumes, high speeds, often in real time. This is also referred to as velocity
SMART ANALYTICS	Uses a range of data analytical tools able to extract information, and glean patterns for a range of uses producing added value where innovative information technologies are combined with evolving mathematical approaches.
AUTOMATION	Uses Machine Learning / Artificial Intelligence techniques where data can be analysed and discernible patterns proposed in real time

Most of the legal, economic and policy literature concerning Big Data uses functional definitions such as those set out above. The technical community, as expected, has a more refined set of Big Data Taxonomies. For example, one taxonomy proposed by Peter Mell,[13] Chief Scientist with the National Institute of Standards and Technologies, further separates the characteristics of Big Data, associating them with three types of Big Data (see Table 2-3 2-3):

Table 2-3 Big Data characteristics and derivation of a notional taxonomy

Volume	Velocity	Variety (semi-structured or unstructured)	Requires Horizontal Scalability	Relational Limitation	Big Data
No	No	No	No	No	Other / Not Big Data
No	No	Yes	No	Yes	Type 1
No	Yes	No	Yes	Maybe	Type 2
No	Yes	Yes	Yes	Yes	Type 3
Yes	No	No	Yes	Maybe	Type 2
Yes	No	Yes	Yes	Yes	Type 3
Yes	Yes	No	Yes	Maybe	Type 2
Yes	Yes	Yes	Yes	Yes	Type 3

Type 1: This is where a non-relational data representation is required for effective analysis.
Type 2: This is where horizontal scalability is required for efficient processing.
Type 3: This is where a non-relational data representation processed with a horizontally scalable solution is required for both effective analysis and efficient processing.

[13] Peter Mell, 'NIST Presentation: Overview of Big Data and Security Implications', National Institute of Standards and Technology, 2015 <http://breakinggov.sites.breakingmedia.com/wp-content/uploads/sites/4/2012/11/bigdata.pdf>.

Why are taxonomies important? Most of the legal and policy commentary on Big Data does not account for notional differences including: different processing procedures or levels of volume, velocity and variety (use of data that is not structured). Horizontal scalability refers to a cost effective mechanism to examine large datasets where the data is broken into smaller sets that are distributed over multiple servers. Relational limitations refer to the ability to scale data. With Big Data scalability is essential. Most regulatory and policy analysis is focused on the type of data (financial, metadata, etc.) or the organisation that owns or is responsible for the data (Tax Office, Bureau of Statistics, etc.).[14] The omission of notional and conceptual Big Data taxonomies in regulatory and policy analysis can lead to over-generalisation or inadequate assessment of risks. Often this equates to economic risk where money is invested to use Big Data but the system isn't scalable or the data feeding into the system isn't done in a manner that will bring about efficiencies. Big Data suffers from the lack of an accepted taxonomy, which is particularly problematic in interdisciplinary discussions. For a data scientist who understands horizontal scalability and the importance of relational versus non-relational data, there is a general consensus as to what Big Data is and is not. For others, as will be seen from Stakeholders' Perspectives chapters in the country reports, Big Data simply means having a lot of data such an online digital library, or having improved ability to process the data.

2.2 Perceived Benefits of Big Data

There are many benefits and risks in using Big Data for law enforcement and intelligence, as well as for national security purposes. While some of these benefits and risks are specific to the national security context, most are equally applicable to commercial and other governmental purposes. As Big Data analytics is evolving both in its use and through technical advancements, the perceived benefits and risks of Big Data may not eventuate as predicted. Further, some impacts may be perceived as both benefits and risks by different stakeholders. These overlapping impacts are: privacy, transparency, de-identification and re-identification, security, and data control.

We also asked research participants to describe what they saw as the capabilities, opportunities and possibilities associated with Big Data (see country reports). While many of the benefits below are captured in participant responses, some have more prominence than others.

[14] Ben Grubb, 'Data retention discussion shrouded in secrecy', *Technology, The Sydney Morning Herald* (Sydney), 26 August 2014 <http://www.smh.com.au/digital-life/digital-life-news/data-retention-discussion-shrouded-in-secrecy-20140826-108fdr.html>; UN General Assembly, Resolution on 'The Right to Privacy in the Digital Age,' Document A/RES/68/167, 18 December 2013 <http://www.un.org/ga/search/view_doc.asp?symbol=A/68/456/Add.2> or <http://justsecurity.org/wp-content/uploads/2014/03/GA-Resolution-Privacy-in-the-Digital-Age.pdf; Alana Maurushat, Lyria Bennett-Moses and David Vaile, 'Using "Big" Metadata for Criminal Intelligence: Understanding Limitations and Appropriate Safeguards,' (June, 2015) Proceedings of the 15th International Conference on Artificial Intelligence and Law 196 <http://dl.acm.org/ft_gateway.cfm?id=2746110&ftid=1596343&dwn=1&CFID=542654113&CFTOKEN=24738647>; Parliamentary Joint Committee on Law Enforcement, 'Inquiry into financial related crime', Report, 7 (September 2015) <http://www.aph.gov.au/Parliamentary_Business/Committees/Joint/Law_Enforcement/Financial_related_crime/Report> .

2.2.1 Smart Analytics

Smarter analytics is promoted as the greatest potential benefit of using Big Data and Big Data tools. 'Smart Analytics' is an umbrella term suggesting a qualitative difference between earlier data mining techniques.[15] Smarter analytics allows for improved predictive analysis, helps detect previously unknown patterns for analysis capabilities, produces results in new forms of visualisation, and produces scalable, integrated high performing analysis.[16] While some of the above has been perceived as marketing hype[17], it is difficult to ignore the many success stories from corporations and government organisations resulting from improved analytics.[18]

Machine learning

Machine learning developed out of the fields of pattern recognition and artificial intelligence.[19] Machine learning algorithms 'learn' from data to identify correlations and patterns, often in order to make predictions.

Many Big Data tools have been developed for law enforcement, with different nations opting for different controls around both the data and decisions made based on the data.[20] For example, France applies strict control on Big Data tools and requires personal control over the data collected and requires human involvement in decisions made based on that data. Spain is more lenient, allowing automated machine learning tools to analyse data as well as automated decision-making based on data.[21] There are no standards in Australia specifically relating to Big Data as such.

One advantage of using machine learning is that, in theory, no human eyes would need to see personal information or sensitive data – the algorithm studies the data, determines its utility and relationship with other data and then spits out areas of concern. The algorithm could also de-identify or obfuscate data to reduce privacy concerns. This potentially reduces claims of intrusive privacy invasion, and individual abuse around collection, processing, storage, use and re-use. Automated machine learning coupled with limitations on re-identification can reduce abuse around personal information.

More effective search ranking and task prioritisation

Law enforcement and intelligent analysts receive large quantities of data, sometimes as changing data streams. One of the greatest potential problems in large volumes of data is

---

[15] J Manyika et al, *Big Data: The next frontier for innovation, competition and productivity*, McKinsey Global Institute (online), 2011 <http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation>.

[16] S Morgan and C Winship *Counterfactuals and Causal Inference: Analytical Methods for Social Research* (New York: Cambridge University Press, 2007).

[17] Martha Bennett, 'Opinion: Is Big Data just big hype?' *Computer Weekly* (2 March 2012) <http://www.computerweekly.com/news/2240143590/Opinion-Is-big-data-just-big-hype>

[18] See, for example, CSC, *Big Data* Success *Stories* (various dates) <http://www.csc.com/big_data/success_stories>.

[19] Christopher Bishop, *Pattern Recognition and Machine Learning* (Springer 2006).

[20] Alana Maurushat, Lyria Bennett-Moses and David Vaile, 'Using "Big" Metadata for Criminal Intelligence: Understanding Limitations and Appropriate Safeguards' (Proceedings of ICAIL '15, 15th International Conference on Artificial Intelligence and Law, San Diego USA, 8–12 June 2015) <http://dl.acm.org/citation.cfm?id=2746090&picked=prox>.

[21] P Casanova, 'CAPER Regulatory Model: Platform to Fight Organised Crime' *AustLII Workshop on Privacy* (September 2014) <http://www.austlii.edu.au/austlii/seminars/2014/3.html>.

the perception of 'drowning in data'. Big Data tools can aid the officer or analyst in producing better search rankings for data queries. The analyst then has a tool to assist them with where to prioritise their time and resources. This is particularly important where there are imminent dangerous threats and time is critical.[22]

Perfect personalisation

The term 'perfect personalisation' refers to the ability to integrate many datasets, often processed in real time, and to deliver a service tailored to an identified person. The term has broadly been used by companies delivering targeted marketing specific to an individual. For example, when you shop at a store, products and services use specific advertisements and promotions unique to you, and not to a pre-defined subset of consumers.[23]

Smarter analytics can also present data in ways that are tailored to particular users. For example, some users may be visual learners who prefer analysis to be presented as graphs, while others may be more comfortable with raw data.

Privacy benefits

Much of the existing literature focuses on privacy concerns[24] and issues over user profiling and Big Data analytics.[25] There is, however, mention in the literature of the ability to increase privacy protection through integrated tools.[26] One example of this is where metadata is automatically anonymised and encrypted in order to make re-identification more difficult. Software agents could be layered into the tools to enhance privacy according

---

[22] Vidit Jain and Manik Varma, 'Learning to re-rank: query-dependent image re-ranking using click data' (WWW '11, Proceedings of the 20th international conference on World wide web, 2011) 277-286 <http://dl.acm.org/citation.cfm?id=1963447>; H Chen, RHL Chiang, VC Storey, 'Business Intelligence and Analytics: From Big Data to Big Impact' *MIS quarterly* (2012) <http://hmchen.shidler.hawaii.edu/Chen_big_data_MISQ_2012.pdf>; R Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences* (SAGE, London, 2014).

[23] Liran Einav and Jonathan Levin, 'The Data Revolution and Economic Analysis,' Working Paper No. 19035, National Bureau of Economic Research (2013) in Executive Office of the President of the United States, White Paper, *Big Data: Seizing Opportunities, Preserving Values* (May 2014) 2–3 <http://www.nber.org/papers/w19035>.

[24] For privacy concerns generally in relation to the digital age see the following works: Daniel Solove, 'Introduction: Privacy Self-Management and the Consent Dilemma' (2013) 126 *Harvard Law Review* 1880; Daniel Solove, *Understanding Privacy* (Harvard University Press, 2008); Daniel Solove, 'A Taxonomy of Privacy' (2006) 154 *University of Pennsylvania Law Review* 477, GWU Law School Public Law Research Paper No. 129, January 2006 <http://ssrn.com/abstract=667622>; Daniel Solove and Marc Rotenberg, *Information Privacy Law* (Aspen Publishers, 2003); Graham Greenleaf, 'Privacy in Australia' in James B Rule and Graham Greenleaf (eds), *Global Privacy Protection: The First Generation* (Edward Elgar Publishing, 2008) 141–173.

[25] Francesco Bonchi, and Elena Ferrari, *Privacy-Aware Knowledge Discovery: Novel Applications and New Techniques* (CRC Press, 2010); Julia Lane et al, *Privacy, Big Data and the Public Good: Frameworks for Engagement* (Cambridge University Press, 2014).

[26] K Rannenberg, D Royer and A Deuker (eds), *The future of identity in the information society: challenges and opportunities (*Springer, Berlin, 2009); Michael Froomkin, 'Pets Must Be on a Leash: How U.S. Law (and Industry Practice) Often Undermines and Even Forbids Valuable Privacy Enhancing Technology' (2013) 74(6) *Ohio State Law* (October 1, 2013); Alana Maurushat, Lyria Bennett-Moses and David Vaile, 'Using "Big" Metadata for Criminal Intelligence: Understanding Limitations and Appropriate Safeguards' (ICAIL '15, Proceedings of the 15th International Conference on Artificial Intelligence and Law, June 2015) 196.

to the types of data combined and used.[27] As the probability of re-identification becomes more likely, the underlying algorithm could signal that new protections are required, and automate this process. As such, Big Data tools could be classified as privacy invasive technologies (**PITS**) as well as privacy enhancing technologies (**PETS**). Where these developments are benefits or risks or both is very much dependent on individual perceptions. For discussion of the diverse attitudes to privacy among research participants, see Chapters 2 in each of the country reports.

2.2.3 Improving Predictive Ability

Predicting behaviour

'Predicting Behaviour' is a term that is used interchangeably with predictive analytics and, where relevant, predictive policing. In essence, predictive behaviour involves the use of data analytics (often Big Data analytics) to forecast the behaviour of an individual, organisation or device before it occurs. Put another way, predictive analytics is a set of data tools that refines data mining with raw data into (hopefully) useful information.[28]

These analytical methods promise to provide ready answers to questions such as: how quickly will an epidemic spread and where it will spread; what type of advertisement will entice the user to purchase a product; what is the probability that a parolee will pose a dangerous threat to the community if released from prison; what is the probability that there will be violence at a demonstration; where should police forces be placed to reduce risk of auto theft; who will win an election; and so forth. In each of these cases, quantitative information about correlations and probabilities can be converted into real-world actions through its influence over human decisions.

Predictive techniques based on Big Data tools are already being used in both private and public decision-making and, in particular, by legal practitioners, judges and police.[29] On the private side, software is being developed to predict the outcomes of legal disputes. Lex Machina is a private analytics company founded in 2010 aiming to predict the cost and outcome of intellectual property litigation.[30] A predictive model has also been developed tracking settlement outcomes of securities fraud class action lawsuits.[31] Big Data tools can also be used in electronic discovery to decrease the costs of civil litigation.[32] In the public sector, data analytics has been used in some US jurisdictions to make decisions about bail

---

[27] V Garg, S Patil, A Kapadia, and LJ Camps, 'Peer-Produced Privacy Protection,' (2013) *Technology and Society* (ISTAS), 2013 IEEE International Symposium on Privacy; Hafeez Manuwar, 'A pattern language for developing privacy enhancing technologies' (2013) 43(7) *Software: Practice and Experience*.

[28] E Siegel, *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die* (John Wiley & Sons, 2012).

[29] Lyria Bennett Moses and Janet Chan, 'Using Big Data for Legal and Law Enforcement Decisions: Testing the New Tools,' (2014) 37(2) *UNSW Law Journal* 643.

[30] Lex Machina, *About Us* <https://lexmachina.com/about/>.

[31] Blakeley B McShane et al, 'Predicting Securities Fraud Settlements and Amounts: A Hierarchical Bayesian Model of Federal Securities Class Action Lawsuits' (2012) 9 *Journal of Empirical Legal Studies* 482.

[32] Nicholas Pace and Laura Zakaras, *Where the Money Goes: Understanding Litigant Expenditures for Producing Electronic Discovery* (RAND Institute for Civil Justice, 2012) 62–6 <http://www.rand.org/content/dam/rand/pubs/monographs/2012/RAND_MG1208.pdf>; Tonia Hap Murphy, 'Mandating Use of Predictive Coding in Electronic Discovery: An Ill-Advised Judicial Intrusion' (2013) 50 *American Business Law Journal* 609.

based on an 'an objective, scientific measure of risk'.[33] A recent Arnold Foundation report advocated growth of the less than 10 per cent of US jurisdictions using data analytic tools in pre-trial decision-making, arguing this tool should be available to all judges as an aid to pre-trial decision-making in order to 'make our communities safer and stronger, our corrections budgets smaller, and our system fairer.'[34] Big Data analytics may also be relevant in post-conviction decisions. Some jurisdictions, such as Virginia, link parole decisions to statistical data concerning rates of reoffending for people in different categories.[35] It is based on a points system, which counts how many factors, statistically aligned with reoffending rates, are present in a particular case. Some of the factors are intuitive and likely to be relevant even in the absence of data. Others, such as the gender of the victim in sex offences, are both less intuitive and more problematic.

Predictive policing

Predictive policing is also being explored in some jurisdictions as a means of optimising police deployments to match predictions about *who* will commit crimes and *where*.[36] Previously undetected patterns may become critical aspects for further investigation which may lead to the identification of suspects and targets.[37] These uses are currently relatively small-scale, and mainly confined to the US, although Australia is beginning to recognise the potential of Big Data analytics, with growing interest and investment in research in that field.

Predictive policing is a form of predictive behaviour specifically within the landscape of law enforcement and intelligence. It has been defined as:

> … a multi-disciplinary, law enforcement-based strategy that brings together advanced technologies, criminological theory, predictive analysis, and tactical operations that ultimately lead to results and outcomes -- crime reduction, management efficiency, and safer communities. Predictive policing builds on concepts from community policing and problem solving. It enhances and expands comprehensive computer statistics (**Compstat**) for accountability purposes and crime reduction. It also makes use of established and long-known 'predictor variables' developed from criminological research.[38]

Big Data tools and analytics are used in a variety of ways in predictive policing. Social media can be analysed in real-time and factored into other datasets to predict the probability of protests or vandalism post sporting events. Historical crime mapping can be used to not only

---

[33] Anne Milgram, *Why Smart Statistics Are the Key to Fighting Crime* (October 2013) TED 8:48 <http://www.ted.com/talks/anne_milgram_why_smart_statistics_are_the_key_to_fighting_crime/transcript#t-27602>.

[34] Laura and John Arnold Foundation, 'Developing a National Model for Pre-trial Risk Assessment '(November 2013) 5 <http://arnoldfoundation.org/sites/default/files/pdf/LJAF-research-summary_PSA-Court_4_1.pdf>.

[35] Bernard Harcourt, *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age* (University of Chicago Press, 2007) 13–14.

[36] Craig D Uchida, 'Predictive Policing' in Gerben Bruinsma and David Weisburd (eds), *Encyclopedia of Criminology and Criminal Justice* (Springer, 2013) 3871, 3871.

[37] P Alston, 'CIA and Targeted Killings beyond Borders' (2011) 2 *Harvard National Security Journal* 283 <http://harvardnsj.org/wp-content/uploads/2011/02/Vol.-2_Alston1.pdf>; BE Harcourt, *Against Prediction: Profiling, Policing and Punishing in an Actuarial Age* (University of Chicago Press, 2007).

[38] Craig Uchida, 'Predictive policing in Los Angeles: Planning and development' *Justice and Security Strategies* (2013) <http://newweb.jssinc.org/wp-content/uploads/2012/01/Predictive-Policing-in-Los-Angeles.pdf>.

predict the likelihood of certain acts such auto theft but then combine historical data with current real-time information such as weather, and information on lighting in car parks to better predict the likelihood of a criminal act occurring in specific locations and at certain times of day.[39] Predictive policing is linked to evidence based policy, real-time evidence, preventative policing and mitigation of unwanted and dangerous events, and effective and efficient use of resources.

Predictive policing has gone beyond clever analysis of Compstat and into a new era of mathematical forecasting and machine learning, or put another way, algorithmic criminology with the use of random forests, stochastic gradient boosting, and Bayesian additive trees.[40]

Predicting behaviour (inclusive of predictive policing and predictive analytics) represents one of the most promising applications of Big Data, and is one of two aspects most written about in the literature.[41] The other is privacy.

Predicting imminent dangers

Predicting behaviour and predictive policing go beyond the ability to forecast crime.[42] Big Data is tied in with smarter analytics, real-time ability, and task prioritisation. This leads to a number of situations where imminent dangers become not only predictable, but the danger is forecast earlier.[43] This may allow agencies timely knowledge to prevent and avoid imminent danger, better ability to mitigate undesirable consequences resulting from the danger, and ability to better predict and avoid subsequent events triggered from the danger or threat.

These types of dangers are often written up in the national security space as avoiding terrorist attacks and other terror related incidents.[44] Less known are the national security threats outside of terrorists plotting bombs, attacks and murders. Big Data tools and analytics are also used in the national security space to:

---

[39] WL Perry et al, 'Predicting Policing: The Role of Crime Forecasting', in *Law Enforcement Operations* (Rand Corporation, 2013) <http://www.rand.org/pubs/research_reports/RR233.html>.
ER Groff and NG La Vigne, 'Forecasting the future of predictive crime mapping' in N Tilley (ed) *Analysis for Crime Prevention* (Monsey, NY: Criminal Justice Press and Devon: Willan, 2002) 29–57.
[40] R Berk, 'Algorithmic criminology' (2013) *Security Informatics* 2:5 <http://www.security-informatics.com/content/2/1/5>; R Berk and J Bleich, 'Statistical procedures for forecasting criminal behavior' (2013) 12(3) *Criminology & Public Policy* 513 <http://www-stat.wharton.upenn.edu/~berkr/Bake-Off%20copy.pdf>.
[41] Janet Chan and Bennett-Moses, 'Is Big Data Challenging criminology?' (2015)*Theoretical Criminology* 1; MS Gerber, 'Predicting crime using Twitter and kernel density estimation' *Decision Support Systems* 61, 115; BE Harcourt, *Against Prediction: Profiling, Policing and Punishing in an Actuarial Age* (University of Chicago Press, 2007); S Morgan and C Winship, *Counterfactuals and Causal Inference: Analytical Methods for Social Research* (Cambridge University Press, New York, 2007); WL Perry et al, 'Predicting Policing: The Role of Crime Forecasting,' in *Law Enforcement Operations* (Rand Corporation, 2013) <http://www.rand.org/pubs/research_reports/RR233.html>; Craig Uchida, 'Predictive policing' in G Bruinsma and D Weisburd (eds) *Encyclopedia of Criminology and Criminal Justice* (Springer, New York, 2013) 3871–3880.
[42] R Berk, 'Algorithmic criminology', *Security Informatics* 2:5 <http://www.security-informatics.com/content/2/1/5>.
[43] Alexander Olesker, '*White Paper: Big Data Solutions for Law Enforcement*', white paper, CTOlabs (June 2012) <http://ctolabs.com/wp-content/uploads/2012/06/120627HadoopForLawEnforcement.pdf>.
[44] Babak Akhgar et al, *Application of Big Data for National Security A Practitioner's Guide to Emerging Technologies* (Elsevier, 2015).

- predict vulnerabilities in critical infrastructure such as water systems, electrical grids, and telecommunications;
- forecast election results (critical to deployment of troops to violent prone regions of the world);
- understand immigration patterns (only very few instances of immigration warrant the guise of national security);
- detect pandemics; and
- respond to natural disasters[45]

While Big Data analytics may be beneficial in a wide array of prediction and mitigation against dangers or threats, this does not mean that traditional methods of law enforcement and intelligence should be abandoned, but merely that the focus has changed from post-event to pre-event. Events could be predicted, possibly prevented and/or disrupted, as opposed to merely reacting post-event.

2.2.4 Supporting Greater Operational Efficiency

Evidence based use

Governments are increasingly concerned with evidence based decision-making. Australia has pushed an evidence-based policy framework for the past decade. For instance, there was a Productivity Commission roundtable on the topic of strengthening evidence-based policy making.[46] Evidence-based policy has become the intended norm in decision-making. Traditionally, evidence-based policy making has relied on empirical studies, surveys, and stakeholder workshops. In the era of Big Data, evidence-based policy making is potentially transformed into something much greater. Lawmakers are now able to supplement traditional studies with real-time analysis of issues and events. For example, looking at past studies conducted over a decade analysing when violence occurs during or after sporting events would be useful to law enforcement for prevention. One can imagine using smart analytics monitoring a sporting event in real-time, predicting the likelihood of violence, the likely location of violence within a stadium, and the likelihood of vandalism in different adjacent neighbourhoods.

Evidence provided in real time

Real-time surveillance and evidence collection is not a new concept; Australian Internet Service Providers (**ISPs**) have had real-time data collection and analytics capability since the mid-2000s.[47] Real-time data collection is also referred to as real-time forensics or live forensics, and differs from post-mortem forensics and data capture. Real-time data collection allows information that could consist of running processes, event logs, network information, protocols, running services (programs and protocols used on a device), and

---

[45] Fleur Johns, 'Data Mining as Global Governance' in *The Oxford Handbook on the Law and Regulation of Technology,* edited by Roger Brownsword, Eloise Scotford and Karen Yeung (Oxford University Press, forthcoming 2016).

[46] Productivity Commission, *Strengthening Evidence-based Policy in the Australian Federation* Australian Government (2012) <http://www.pc.gov.au/research/completed/strengthening-evidence>. But see Lyria Bennett Moses, Kieran Tranter and Nicola Gollan, 'The Productivity Commission: a different engine for law reform?' (2015) *Griffith Law Review* (pre-published online) [2015] *UNSWLRS* 40 <http://www.austlii.edu.au/au/journals/UNSWLRS/2015/40.html>.

[47] Productivity Commission, *Strengthening Evidence-based Policy in the Australian Federation*, Australian Government (2012) <http://www.pc.gov.au/research/completed/strengthening-evidence>.

metadata. Past use of real-time evidence collection involved data to be stored. Where real-time data is stored, law enforcement agents are potentially able to view emails pre-crime, post-crime and during the commission of a crime.

The legal regime concerning access to telecommunications metadata, and the ability to integrate such data into other data sets, is complex. If it can be accessed in bulk and combined with other data sets, it enhances the ability of law enforcement to track communications in real time during critical incidents.

More effective and efficient use of resources

In an era of government cutbacks, and emphasis on combining departments and reducing costs, Big Data offers the potential to use resources more efficiently.[48] While the initial cost to invest in Big Data tools and resources may be high, the long term benefit of having smarter analytics allows prioritisation of duties, reduction in time taken to analyse data, and in the long term, the reduction of expenses. In addition to long term cost reduction, smart analytics should enhance the ability of government to direct resources to areas most in demand. This strategic alignment of resources can assist both in short and long term resource management.[49]

Security

Security is often said to go hand in hand with privacy. Like privacy, security is also a potential benefit and risk. Depending on the underpinnings of an algorithm, Big Data tools can be used to enhance security or can instead pose a security risk.[50] Again this is dependent on how the technologies evolve, and whose perception is being considered.[51]

There are a variety of ways in which Big Data can be used to enhance security. They can be used for fraud detection (including the detection of unauthorised access to data) as well as in security incident and event management (**SIEM**). The use of Big Data systems could potentially enhance intrusion detection systems (**IDS**) and intrusion prevention systems (**IPS**) as the systems are machine-learning allowing for continual adjustment and update changes. Big Data methods potentially provide a means to reduce silos while preserving some privacy and security enhancing features. These techniques could also help to detect threats at an earlier stage.[52]

---

[48] Council of Australian Governments (COAG), *Final Report of the COAG Review of Counter-Terrorism Legislation*, 1 March 2013 <http://www.ag.gov.au/Consultations/Pages/COAGReviewofCounter-TerrorismLegislation.aspx>.

[49] P Casanova, 'CAPER Regulatory Model: Platform to Fight Organised Crime, *AustLII Workshop on Privacy* (September 2014) <http://www.austlii.edu.au/austlii/seminars/2014/3.html>.

[50] Lei Xu, 'Information Security in Big Data: Privacy and Data Mining' (2014) 2 *IEEE Access* 1149 <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6919256&newsearch=true&queryText=Information%20Security%20in%20Big%20Data:%20Privacy%20and%20Data%20Mining>.

[51] Babak Akhgar et al, *Application of Big Data for National Security: A Practitioner's Guide to Emerging Technologies* (Elsevier, 2015); Jane Bambauer, Krish Muralidhar, and Rathindra Sarathy, 'Fool's Gold: an Illustrated Critique of Differential Privacy' (2014) 16 *Vanderbilt Journal of Entertainment & Technology Law*; Laura Donohue, 'High Technology, Consumer Privacy, and US National Security' (2015) *Georgetown Law Faculty Publications*, 1457; Sean Richmond, 'National security debate misses big picture of 'balanced' response', *The Conversation* (online), 25 February 2015 <https://theconversation.com/national-security-debate-misses-big-picture-of-balanced-response-37923>.

[52] Peter Wood, 'How to Tackle Big Data From a Security Point of View', *Computer Weekly*, March 2013 <http://www.computerweekly.com/feature/How-to-tackle-big-data-from-a-security-point-of-view>.

Ability to better monitor abuse or unauthorised access of data

Because Big Data involves smarter analytics and larger datasets, methods have been fine tuned to better control who accesses and uses data, under conditions, and for what purpose.[53] This can take many forms. Big Data tools are said to be able to better monitor unauthorised access of data, unauthorised usage, and unauthorised distribution. This in return is said to aid in decreasing abuse of data – whether the abuse is accidental or deliberate.[54] Control of data is linked to improved privacy and security of data.

Provenance

Provenance is a term to describe the origin of something, or put differently the metadata about data. Data provenance is information about the creation or origin of data and data processes. Data provenance has been studied extensively in data fields such as modelling, database management, and distributed systems.[55] The study of data provenance in Big Data is recent.[56] The technical literature discusses how provenance may be used to improve system performance, analysis, produce metrics and test for exploits and bugs.[57]

The implications of data provenance as applied to legal issues and policy decisions are not readily referred to in the literature.[58] Provenance in this sense would refer to the inferences, reasons and weighting behind choices in system output that are identifiable by analysts, enhancing reliability and accountability of subsequent decision-making. Maintaining provenance would enhance comprehensibility and thus the accountability of decisions relying on inferences drawn.

---

[53] Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press, 2015).

[54] Council of Australian Governments (**COAG**), *Final Report of the COAG Review of Counter-Terrorism Legislation*, 1 March 2013 <http://www.ag.gov.au/Consultations/Pages/COAGReviewofCounter-TerrorismLegislation.aspx> or <http://www.ag.gov.au/Consultations/Documents/COAGCTReview/Final%20Report.PDF>; Robert Bloom and William Dunn, 'The Constitutional Infirmity of Warrantless NSA Surveillance: The Abuse of Presidential Power and the Injury to the Fourth Amendment' (2007) 15 *William and Mary Bill of Rights Journal* 147 <http://lawdigitalcommons.bc.edu/lsfp/163/>; Roger Clarke, '*Privacy Impact Assessments as a Control Mechanism for Australian National Security Initiatives,'* Xamax Consultancy (2015) <http://www.rogerclarke.com/DV/IANS.html>.

[55] Olaf Hartig, 'Provenance Information in the Web of Data' (2009) Linked Data on the Web 1 <http://events.linkeddata.org/ldow2009/papers/ldow2009_paper18.pdf>; Peter Buneman and Susan Davidson, 'Data provenance – the foundation of data quality' in *Data provenance – the foundation of data quality* (2013) <http://www.sei.cmu.edu/measurement/research/upload/Davidson.pdf>.

[56] Rajeev Agrawal et al, 'A layer based architecture for provenance in Big Data' (2014) *IEEE International Conference on Big Data* 29. <http://ieeexplore.iee.org/xpls/abs_all.jsp?arnumber=7004483&tag=1>; Gustavo Alonso, Boris Glavic, 'Perm: Processing Provenance and Data on the Same Data Model through Query Rewriting' (2009) *IEEE 25th International Conference on Data Engineering* 174 <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=4812401&tag=1>; Carl Lagoze, 'Big Data, data integrity, and the fracturing of the control zone' (2014) 1 *Big Data & Society* 1; <http://bds.sagepub.com/content/1/2/2053951714558281>.

[57] Boris Glavic, 'Big Data provenance: challenges and implications for benchmarking' in Tilmann Rabl et al, *Specifying Big Data Benchmarks* (Springer Berlin Heidelberg 2014) 72 <http://link.springer.com/chapter/10.1007/978-3-642-53974-9_7#>.

[58] Chan J and Bennett Moses L, 'Is Big Data Challenging criminology?' (2015)*Theoretical Criminology*, 1; S Morgan and C Winship, *Counterfactuals and Causal Inference: Analytical Methods for Social Research* (Cambridge University Press, New York, 2007).

Reverse provenance refers to a situation where an analyst can deduce whether a particular prediction (sourced from other intelligence) could have been predicted from data such as open source social media data. It is useful in preserving operational secrecy in that it allows agencies to deduce whether information from a particular source was discoverable without that source.

2.3 Perceived Risks of Big Data

2.3.1 Key Challenges

Drowning in data

Some users of datasets and data analytics are concerned that the volumes involved in Big Data will lead to significantly more data to analyse with minimal gain in benefits, hence the term 'drowning in data'.[59] While this remains a concern amongst data analysts this risk is mostly absent from the literature, or is reflected within the context of the Snowden leak of NSA documents, and speculation as to whether the extensive surveillance has proven an efficient method of preventing terrorism and preventing national security.[60]

Shortage of trained professionals in data analytics

There is a reported global shortage of skilled data scientists of the level required for the implementation of Big Data systems. The shortage is further amplified by those with the relevant skills accepting higher salaries within the private sector. While not a new issue, the level of skill required for high quality data science has exacerbated the problem. The other problem is that there is a lack of training for decision-makers on the assumptions and methods lying behind inferences drawn from Big Data.

Use of local cultural patterns as basis for interpreting foreign behaviours

While governments are increasingly moving towards evidence based policy-making, statistics and data analytics still rely heavily on trained personnel to make decisions based on the data. When data analysis identifies a pattern, it is important to not make wider assumptions about these patterns beyond their scope. The literature specifically calls for caution when interpreting patterns that may be accurate, but only from a local cultural perspective.[61] The inverse is also true – when agencies look at inferences drawn from Big Data by foreign counterparts, they need to recognise that these patterns may not be accurate in a local cultural context.[62]

---

[59] Adrian Lawrence and L Lim, *Privacy and Security: Introduction – Privacy in Cyberspace,* Law of E-Commerce (2015 Lexis Nexis).

[60] Henry Farrell and Martha Finnemore 'The End of Hypocrisy: American Foreign Policy in the Age of Leaks' (2013) *Foreign Affairs* 22 (Nov/Dec 2013); Kit O'Connell 'Why The Media Ignores Jeremy Hammond While Praising Edward Snowden', MintPress News (online), 11 May 2015 <http://www.mintpressnews.com/why-the-media-ignores-jeremy-hammond-while-praising-edward-snowden/205501/>; John Yoo, 'The Legality of the National Security Agency's Bulk Data Surveillance Programs', 10 *ISJLP* 301 (2014) <http://scholarship.law.berkeley.edu/facpubs/2429>.

[61] d boyd and K Crawford, 'Critical questions for Big Data: Provocations for a cultural, technological and scholarly phenomenon' (2012)15(5) *Information, Communication and Society* 662.

[62] Kate Westmore and Gail Kent*, International Law Enforcement Access to User Data: A Survival Guide and Call for Action* (8 January 2015) <http://ssrn.com/abstract=2547289> or <http://dx.doi.org/10.2139/ssrn.2547289>.

Reliability and Accuracy of Data

There has been some concern that a move to Big Data systems will undercut the reliability and accuracy of data within datasets. By and large, however, the literature merely states that reliability and accuracy may be a minor risk with a paucity of literature reflecting on decreased accuracy in Big Data systems.[63] The literature refers to improved reliability and accuracy of inferences drawn from Big Data, so this risk, while present, is one that is presently perceived as limited.[64] However, where there is systemic bias in the data being analysed, the reliability and accuracy of inferences drawn will reduce.

Privacy – uncontrolled sharing and de-identification

Much of the risk literature around Big Data involves privacy and security.[65] This is reflected in both the international literature and the Australian literature.[66] Concerns have been raised as to whether the current privacy framework in virtually all nations is adequate for Big Data.[67] Some have argued that once data has been collected, there is no control over who uses it or how it is used; there is no privacy.[68]

The Canadian Privacy Commissioner called for privacy reforms to accommodate Big Data – Big Data requires Big Privacy.[69] Surprisingly, many technologists and computer scientists have called for greater privacy protection, arguing that the current frameworks are either incompatible with Big Data or require serious amendments.[70] Law and policy scholars are equally sceptical as to the adequacy of privacy law, especially in the era of Big Data and the Internet of Things.[71] Not everyone has called for a new privacy regime, with many merely pointing out the inadequacies in the current regime or calling upon industry to self-regulate.[72]

---

[63] J Laurila et al, 'The Mobile Data Challenge: Big Data for Mobile Computing Research', conference paper EPFL-CONF-192489, *Pervasive Computing* Workshop, Newcastle (2012) <http://infoscience.epfl.ch/record/192489>.

[64] Department of Finance and Deregulation, AGIMO. (2013) *The Australian Public Service Big Data Strategy: Improved understanding through enhanced data-analytics capability*, August 2013 <http://www.finance.gov.au/sites/default/files/Big-Data-Strategy_0.pdf>.

[65] Solon Barocas and Helen Nissenbaum, 'Big Data's End Run around Anonymity and Informed Consent' in Julie Lane et al, *Privacy, Big Data, and the Public Good* (Cambridge University Press, 2014) 44–56.

[66] Marko Milanovic, 'Human Rights Treaties and Foreign Surveillance: Privacy in the Digital Age' (April 2014) *Social Science Research Network* <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2418485>.

[67] Jarrod Bayliss-McCulloch, 'Risks and opportunities in Big Data — how well adapted are Australia's privacy laws?' (2015) 20(1) *Media and Arts Law Review*.

[68] Craig Sebastopol, *Privacy and Big Data: The Players, Regulators, and Stakeholders* (O'Reilly 2011); Erika McCallister, Timothy Grance and Karen Scarfone, 'Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)' *National Institute of Standards and Technology* (2010) <http://csrc.nist.gov/publications/nistpubs/800-122/sp800-122.pdf>.

[69] Anna Cavoukian, and J Jonas, *Privacy by Design in the Age of Big Data* (June 8, 2012) <https://www.ipc.on.ca/images/Resources/pbd-big_data.pdf>.

[70] Ed Arvind Narayanan and Edward W Felten, 'No silver bullet: De-identification still doesn't work' White paper (July 9, 2014) <http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf>.

[71] Des Butler, 'A Tort of Invasion of Privacy in Australia?' (2005) 29 *University of Melbourne Law Review* 339.

[72] Steve Wilson, 'The Collision between Big Data and Privacy Law' (2014) 2(3) *Australian Journal of Telecommunications and the Digital Economy.*

The reference to 'industry' here reflects the fact that most of the privacy literature refers to Big Data use in a commercial context. The literature does not adequately consider privacy issues relating specifically to law enforcement and intelligence use of Big Data, particularly in Australia.

There are safeguards which could reduce privacy concerns and discriminatory profiling. De-identification and anonymisation of data are limited safeguards.[73] For our purpose, 'de-identification' is the removal, stripping or obfuscation of directly identifying elements from a data record or set, such that the result is not immediately identifiable as associated or linked with a particular individual.[74] 'Anonymisation' is the process of rendering data into a form that does not identify individuals in circumstances where identification is not likely to take place. [75] The main difference between de-identification and anonymisation is that de-identified data may potentially be re-identified, whereas anonymisation is said to be irreversible. A de-identification technique allows for re-identification,[76] while the law sets the parameters and context when data can be re-identified.[77] With anonymisation, the technology and the regulatory framework forbid re-identification (this may not always prove successful but the goal is to not allow re-identification under any circumstance). The distinction is, however, theoretical as the literature increasingly tells a story of how re-identification is not only possible but probable.[78] Nonetheless the ability to alleviate some privacy concerns through de-identification, pseudonymity and anonymisation of data is prevalent in the literature.

One interesting development is the evaluation of de-identification tools, re-identification tools and the emerging 'best practice' of differential privacy.[79] Differential privacy has been described as 'personal information in a large database that is not modified or released. Instead, a third party, such as a researcher, can submit questions about the information in the database by going through an intermediary piece of software that serves as a privacy guard'.[80] If the threshold risk is too high for privacy, the researcher may choose to abandon the query, or to modify the algorithm to further reduce privacy risk through techniques such as noise where, for example, additional records are added to the system thereby reducing

---

[73] Alana Maurushat, Lyria Bennett-Moses and David Vaile, 'Using "Big" Metadata for Criminal Intelligence: Understanding Limitations and Appropriate Safeguards,' (June, 2015) Proceedings of the 15th International Conference on Artificial Intelligence and Law 196.

[74] Wilson, Petra 'Legal issues of data anonymisation in research,' (2004) 328 *BMJ* 1300; <http://www.bmj.com/content/328/7451/1300>.

[75] UK Information Commissioner's Office, 'Anonymisation: managing data protection risk Code of practice' November, 2012, at <http://ico.org.uk/~/media/documents/library/Data_Protection/Practical_application/anonymisation-codev2.pdf>; Khaled El Emam, *Risky Business: Sharing Health Data While Protecting Privacy* (Trafford, 2013).

[76] Arvind Narayanan and Edward W Felten, 'No silver bullet: De-identification still doesn't work', White paper, 9 July 2014 <http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf>.

[77] Felix Wu, 'Defining Privacy and Utility in Data Sets' <http://lawreview.colorado.edu/wp-content/uploads/2013/11/13.-Wu_710_s.pdf>.

[78] Jane Bambauer, Krish Muralidhar, and Rathindra Sarathy, 'Fool's Gold: an Illustrated Critique of Differential Privacy' (2014) 16 *Vanderbilt Journal of Entertainment & Technology Law*

[79] Daniel Barth-Jones, *Ethical Concerns, Conduct and Public Policy for Re-Identification and De-identification Practice: Part 3 (*Re-Identification Symposium) <http://blogs.law.harvard.edu/billofhealth/2013/10/02/ethical-concerns-conduct-and-public-policy-for-re-identification-and-de-identification-practice-part-3-re-identification-symposium/>.

[80] Anna Cavoukian, and J Jonas, *Privacy by Design in the Age of Big Data*, June 8, 2012 <http://privacybydesign.ca/content/uploads/2012/06/pBig Data- big_data.pdf>.

re-identification risk.[81] Privacy concerns in differential privacy, however, are frequently misunderstood. Observers in these areas often fail to notice that differential privacy does not protect against the inferential predictions that could be made by a nearly omniscient statistician.[82] This leads back to the potential problems with causation and correlation in Big Data, and potential to discriminate which is discussed below.

Discrimination

Computers and the Internet were founded on principles of equality and ethics around non-discrimination.[83] In historical computer science terms non-discrimination has meant that computer code, by its nature of being 0s and 1s cannot discriminate.[84] In this sense it has been stated that data doesn't discriminate, but people making decisions about data and data analytics do. While this could be seen to be true up to a certain extent, there is a growing concern that machine learning Big Data systems could systemically build in correlations that become discriminatory.[85] There are three ways of viewing concerns about discrimination. The first relates to false inferences where the data set on which analytics is performed is biased (perhaps due to racial bias in how criminal activity is recorded in police databases). The second is that data analytics could be discriminatory in its effects.[86] The third, which is related, is that there are potential negative implications of predictive policing and decision-making processes that factor in correlative information derived from even anonymised data sets. For example, an individual may be affected, or a group stigmatized, because they live or operate in a place or a manner identified through algorithms as correlated with criminal activities.[87]

2.3.2 Creating a Surveillance Society

Escalation of cross-border surveillance

Nations have always formed alliances and shared information in limited fashions. Big Data systems are being developed for intelligence and law enforcement purposes globally. The question becomes, if one nation is doing something, by necessity does another nation need to keep pace? With leaks to the media and online leak sites detailing Big Data capacities of governments and organisations, there is a risk that the adoption of Big Data will proliferate

---

[81] Cynthia Dwork, 'Differential Privacy' *33rd International Colloquium on Automata Languages and Programming (ICALP), Lectures in Computer Science* (2006) http://dx.doi.org/10.1007/11787006_1; K El Eman, *Guide to the De-Identification of Personal Health Information* (CRC Press 2013) <http://www.crcpress.com/product/isbn/9781466579064>.

[82] Daniel Barth-Jones, *Ethical Concerns, Conduct and Public Policy for Re-Identification and De-identification Practice: Part 3 (*Re-Identification Symposium) <http://blogs.law.harvard.edu/billofhealth/2013/10/02/ethical-concerns-conduct-and-public-policy-for-re-identification-and-de-identification-practice-part-3-re-identification-symposium/>.

[83] Computer Ethics, *Wikipedia* <https://en.wikipedia.org/wiki/Computer_ethics>.

[84] Pekka Hineman, *The Hacker Ethic* (Random House 2010); Eric Raymond, *The Cathedral and Bazaar* (O'Reilly 1999).

[85] Mark Burdon and Paul Harpur, 'Re-conceptualising Privacy and Discrimination in the Age of Talent Analytics' *UNSW Law Journal* 37(2) (2014); S Morgan and C Winship, *Counterfactuals and Causal Inference: Analytical Methods for Social Research* (New York: Cambridge University Press, 2007).

[86] Eg Solon Barocas and Andrew D Selbst, 'Big Data's Disparate Impact', *California Law Review* (forthcoming 2016).

[87] Centre for Information Policy Leadership, *'Big Data: A Tool for Inclusion or Exclusion?'* (October 2014), <http://www.informationpolicycentre.com/files/Uploads/Documents/Centre/CENTRE_Big_Data_Workshop_Project_P145406.pdf>.

promoting an escalation of cross-border surveillance.[88] In other words, the more data *we* collect about *them*, the more data *they* collect about *us*. This is not a legal risk, but a risk of a political nature.[89]

Unregulated public partnerships using commercial data

Sharing of information between certain government agencies is highly regulated for many reasons including privacy, corruption, and security. For example, the Australian Tax Office is not permitted to disclose information about tax returns to the Transportation Minister or any other entity not otherwise stipulated under its enabling legislation.

However, private entities sharing data with government agencies are largely unregulated. Private corporations are free to share information with government agencies provided that they comply with privacy law.[90] As long as the user is informed about the privacy terms in the agreement, and the data is shared under the conditions of the agreement (often vague), then data may move freely from private actors to the government. The primary exception is stored communications which cannot be intercepted or accessed absent a warrant from law enforcement. These issues have received little attention in the literature.

Freedom of expression and other civil liberties

Privacy is often seen as the parcel that reveals all of the other human rights and civil liberties inside.[91] Privacy protection enables freedom of expression, freedom of association, freedom of religion and other human rights and civil liberties[92]. These freedoms and civil liberties are considered tenets of Western democracies. When Big Data systems insufficiently balance privacy with security, this impacts on human rights and democracies in general.[93]

Fear of living in a mass surveillance society

Mass surveillance was first introduced into our minds through literature commencing in the early 1900s with Russian physicist, Yevgeny Zamatin, in the book *WE*. Orwell used many aspects of *WE* to later write the famous book *1984*. Since then, many other books, movies and television shows have addressed the dangers of living in a mass surveillance society.

---

[88] P Alston, 'CIA and Targeted Killings beyond Borders' (2005) 2 *Harvard National Security Journal* 283 <http://harvardnsj.org/wp-content/uploads/2011/02/Vol.-2_Alston1.pdf>; Allie Coyne, 'AFP reports data sharing with Russia, China', (18 June 2015) *iTnews* <http://www.itnews.com.au/News/405403,afp-reports-data-sharing-with-russia-china.aspx>.

[89]Ioanna Tourkochoriti, 'The Transatlantic Flow of Data and the National Security Exception in the European Data Privacy Regulation: In Search for Legal Protection Against Surveillance', (2015) 36 *U Pa J Int'l L* 459 <http://scholarship.law.upenn.edu/jil/vol36/iss2/3>.

[90] Joel Reidenberg 'International Approaches to Public and Private Sector Data Privacy and Security' in Peter M Shane, John Podesta and Richard C Leone *A Little Knowledge: Privacy, Security, and Public Information after September 11* (Century Foundation Press, 2004) 98–9.

[91] Etzioni, Amitai, 'NSA: National Security vs. Individual Rights' (2015) 30(1) *Intelligence and National Security* 100.

[92] Marko Milanovic 'Human Rights Treaties and Foreign Surveillance: Privacy in the Digital Age' in his paper on the Social Science Research Network (March 31 2014) <http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2418485>.

[93] Jane Bambauer, 'Other Peoples' Papers' (2015) 94 *Tex. L Rev*; Rice, Simon, ANU Brandis receives long list of rights-limiting laws – now can he justify them? The Conversation, 6 August 2015 <https://theconversation.com/brandis-receives-long-list-of-rights-limiting-laws-now-can-he-justify-them-45645>; Ian Brown et al, *Towards Multilateral Standard for Surveillance Reform*, January 2015, Centre for the Internet and Human Rights <https://cihr.eu/wp-content/uploads/2015/01/Brown_et_al_Towards_Multilateral_2015.pdf>.

People continue to fear the move towards mass surveillance as evidenced in the wider fictional media, as well as in current political and media debates around technical capacity. There is the risk that any forms of Big Data expansion will be conceived as Orwellian and therefore undesirable.[94]

Erosion of Trust

The erosion of trust – whether factual or perceived – is a central issue in the literature. While there are many different facets to trust with law enforcement and intelligence, in the context of Big Data trust issues manifest in two main areas. The first is a grave concern around mass surveillance.[95] The second is erosion of trust of the community with law enforcement if face-to-face community policing is replaced or reduced by relying and spending more time and resources on data analytics.[96]

2.3.4 Challenging Accountabilities

Accountability

The literature is focused on different types of accountability issues. The first relates to the inadequacy of privacy and security legal frameworks for the exploitation of Big Data by business (as already explored above). The second set of literature relates more generally to making intelligence agencies actions more visible and accountable.[97] This literature suggests that current accountability and oversight mechanisms have been reduced over time, and are inadequate for intelligence practices.[98] There is less critique within the literature of accountability within law enforcement compared with intelligence. The third set of literature focuses on the lack of publicly available impact assessment tests.[99] The fourth refers to general accountability and data governance structures appropriate for the use of Big Data.[100] The final set of literature is concerned more specifically with the metadata debate around

---

[94] Chalmers, Robert, 'Orwell or all well? The rise of surveillance culture' (2005) 30(6) *Alt LJ* 258 <http://www5.austlii.edu.au/au/journals/AltLawJl/2005/77.pdf>.

[95] Robert Bloom and William Dunn. 'The Constitutional Infirmity of Warrantless NSA Surveillance: The Abuse of Presidential Power and the Injury to the Fourth Amendment.' *William and Mary Bill of Rights Journal* 15, (2007): 147–202 <http://lawdigitalcommons.bc.edu/lsfp/163/>.

[96] Judy Putt (ed), 'Community Policing in Australia', *The Australian Institute of Criminology* (2010) <http://www.aic.gov.au/media_library/publications/rpp/111/rpp111.pdf>; Sean Richmond, 'National security debate misses big picture of 'balanced' response', *The Conversation* (25 February 2015) <https://theconversation.com/national-security-debate-misses-big-picture-of-balanced-response-37923>.

[97] Hal Abelson, Ken Ledeen and Harry Lewis *Blown to Bits: Your Life, Liberty and Happiness After the Digital Explosion* (Addison-Wesley 2008) 48–55; Bruce Schneier, *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World* (W W Norton & Company 2015) 74–6.

[98] Australian and NSW Government, *Martin Place Siege Joint Commonwealth – New South Wales Review*, (January 2015); David Cole, 'We Are All Foreigners: NSA Spying and the Rights of Others' (October 2013) *Just Security Online* <http://justsecurity.org/2668/foreigners-nsa-spying-rights/>.

[99] ABC TV, 'New security laws being introduced without proper impact assessments,' *Lateline* (25 August 2015)<http://www.abc.net.au/lateline/content/2015/s4299342.htm; Roger Clarke, Approaches to Impact Assessment (2014) <http://www.rogerclarke.com/SOS/IA-1401.html>.

[100] Department of Finance and Deregulation, AGIMO/ Marc Vickers. (2014) *Draft Guide to Responsible Data Analytics*, 26 June 2014 <http://www.finance.gov.au/node/34900/> and <http://www.finance.gov.au/sites/default/files/Responsible%20Data%20Analytics%20Draft.pdf>.

data retention.[101] Accountability, oversight and transparency are issues in Australia and abroad.

Transparency

Transparency is tied to effectiveness, acceptability and accountability.[102] Transparency is linked to public confidence and trust in law enforcement and intelligence, with even the call for meta-transparency when dealing with open data, Big Data and privacy.[103] There appears to be a consistent call for transparency and in particular data governance structures, procedures around implementation and other aspects of security management for data.[104] The literature is concerned with various aspects of transparency:[105]

- transparency within government as ensuring proper accountability,
- public transparency around access to data-sets,
- public and intra-government transparency around what is done with data, particularly around the nature of analysis performed and the operation of different algorithms employed,
- public transparency around the procedures for how data is managed, including security protocols,
- transparency about how decisions are made when looking at data analytics, at least within government and for individuals affected where natural justice issues arise, and
- transparency as an issue of concern in enabling feedback loops where information is disclosed.[106]

Transparency features extensively in various aspects of the literature with consistent conclusion that it is highly important to the legitimacy of Big Data.

---

[101] Court of Justice of the European Union, Digital Rights Ireland and Seitlinger and Others (Digital Rights Ireland, C-293/12 (2014) OJC 258, 25 February 2012, and Seitlinger, C-594/12 (2014) OJ C 79; Ben Grubb, 'Data retention discussion shrouded in secrecy', *Technology, The Sydney Morning Herald*, (26 August 2014) <http://www.smh.com.au/digital-life/digital-life-news/data-retention-discussion-shrouded-in-secrecy-20140826-108fdr.html>(8 April 2014; Jonathan Mayer et al, 'Evaluating the privacy properties of telephone metadata' (2016) 113(20) *PNAS* 5536–5541 doi:10.1073/pnas.1508081113.

[102] Lyria Bennett Moses, and Janet Chan, 'Using Big Data for Legal and Law Enforcement Decisions: Testing the New Tools,' (2014) 37(2) UNSW Law Journal 643.

[103] Keiron O'Hara, *Transparent government, not transparent citizens: a report on privacy and transparency for the Cabinet Office* (London, GB, Cabinet Office, 2011) 84.

[104] National Audit Office. (2012) *Implementing Transparency* <http://www.nao.org.uk/report/implementing-transparency/>; Attorney-General's Department *Information security management guidelines,* Australian Government (1 November 2014) v2.0 <https://www.protectivesecurity.gov.au/informationsecurity/Documents/AustralianGovernmentclassificationsystem.pdf>.

[105] Danielle Keats Citron and David Gray, 'Addressing the Harm of Total Surveillance: A Reply to Professor Neil Richards' (2013)126 (262) *Harvard Law Review Forum* <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2285775>.

[106] Lyria Bennett Moses and Janet Chan, 'Algorithmic Prediction in Policing: Assumptions, Evaluation, and Accountability,' Data Associations in Global Law and Policy Workshop, UNSW Australia, Sydney, December 2015.

Scope creep

There is a general concern that if a technology or type of data analytic is developed for one purpose, that this purpose could be used for other applications. While this may or may not be a risk (certainly benefits may be derived from this), the concern is more that extension of use of technologies requires detailed and careful consideration.[107] The legal and policy frameworks initiated in the first instance of use may be inappropriate when applied to subsequent uses. Likewise, the public debate (if any) may be limited to the first use, and there may be less transparency and debate around extending the scope.

Understanding Correlation and Causation for Decision Makers

As Big Data often uses multiple large and disparate datasets, it is able to find correlations previously unknown or undiscoverable through other methods. Inferences may be drawn and decisions may be made based on these correlations.[108] However, the distinction between correlation and causation is sometimes poorly understood, particularly among those without data science expertise.[109] In particular, it is important that decision-makers understand that action taken based on inferences drawn from correlations may have an unforeseen impact, for example by generating 'feedback loops'.

Data Quality Assurance

Analytics relies on quality of data. Any inferences drawn from data are dependent on the quality of the data, or at least the absence of systemic bias within the data.[110] Critical data quality factors include accuracy, precision, temporal applicability, currency, completeness, deterioration, reliability, and meaning.

2.3.4 Creating Vulnerabilities

Asset exposure

De-identification and re-identification were considered as benefits previously where suspects could be potentially re-identified based on de-identified data. The risk is that the inverse is also true. Critical assets and targets may also be re-identified risking their safety and security. The literature mostly focuses on these issues in the health information space, but the techniques and consequences of dealing with sensitive data are similar.[111]

There are also concerns that the granularity and control of access to data, in particular who accesses data and under what conditions, may be lost in Big Data systems. In short, agencies have experience with their current systems. They understand the risks and benefits and

---

[107] Anne-Marie Oostveen and Diana Dimitrov, ' scanners can now identify us from 40 feet away', *The Conversation* (21 May 2015) <https://theconversation.com/iris-scanners-can-now-identify-us-from-40-feet-away-42141>.

[108] Harry Surden, 'Machine Learning and the Law' 89 Washington Law Review 1 (2014).

[109] Harcourt, BE (2007) Against Prediction: Profiling, Policing and Punishing in an Actuarial Age. Chicago, IL: University of Chicago Press; Janet Chan and Bennett-Moses, 'Is Big Data Challenging criminology?' (2015)*Theoretical Criminology*, 1.

[110] Roger Clarke, 'Big Data Quality Assurance' (2015) <http://www.rogerclarke.com/EC/BigDataQA.html>; U Fayyad, G Piatetsky-Shapiro and P Smyth (1996) 'From Data Mining to Knowledge Discovery in Databases ' AI Magazine 17, 3 (1996). http://aaaipress.org/ojs/index.php/aimagazine/article/download/1230/1131; Laura Sebastian-Coleman, *Measuring Data Quality for Ongoing Improvement: A Data Quality Assessment Framework* (MK 2013).

[111] Jiuyong Li et al, 'Current Developments of k-Anonymous Data Releasing', *National e-Health Privacy and Security Symposium* (2006) <http://eprints.usq.edu.au/1307/1/13.pdf>.

have grown to trust their systems. There are concerns about the impact of a change to Big Data systems.

Security

It has been said that Big Data requires Big Privacy, and if so, it is presumed that it will likewise require Big Security. There are many security concerns when using and storing large volumes of data. Current data storage systems often classify data by type in order to comply with appropriate legal frameworks. Health information, for example, is considered sensitive personal information and is subject to much stricter privacy and security controls. It has been argued that information classification becomes even more critical when you are drawing from disparate data sources in large volumes. Questions arise about 'ownership' or control over and responsibility for collated data. Further, there are questions about the appropriate security standards for data that combines different types. Encryption may be required for use of some types of data, or when sharing data between agencies. Security issues are compounded by the logistical difficulties of storing encrypting large volumes of data.

There are also issues with the security of Australian government data stored in the cloud, particularly where it is stored in other jurisdictions.

2.3.5 The Risk of Not Using Big Data

Curiously, the literature is silent on what the risks are if law enforcement and intelligence agencies elect not to pursue Big Data. While the 'opportunity costs of not using Big Data' are mentioned in various workshops and symposiums, the literature is generally silent on the consequences of Big Data being embraced by others, without also being employed for law enforcement and intelligence. As Big Data analytics often use open source tools such as Hadoop, it is likely that organised criminal syndicates and terrorist syndicates are making use of Big Data tools. They may even be producing bespoke tools for themselves. The risk is that that nation states that do not use Big Data will be at a disadvantage, both against criminal syndicates and as compared with other nation states.

2.4 Summary

This chapter has reviewed the available international literature on Big Data – its definition, perceived benefits and perceived risks. The next two chapters will describe the objectives and methods of the current research project.

3. RESEARCH QUESTIONS AND OBJECTIVES

3.1 Key Research Questions

The project addresses the following key research questions:

1. What are the current technological applications and future possibilities of Big Data for law enforcement and national security as perceived by designers, users and stakeholders? What are their perceptions of the social, legal or policy implications of Big Data in this context?

2. What are the strategies, policies, laws, regulatory frameworks, practices and technologies relevant to Big Data adopted by Australia, the UK and Canada, and are they perceived as effective in meeting social, legal or policy objectives, including, but not limited to, law enforcement and national security objectives? What responses are needed to deal with potential challenges?

3.2 Research Objectives

The main objectives of the project are:

1. To scope and assess relevant international perspectives, policies, laws and practices that can inform Australian policymakers and support the development of law and best practices; and

2. To develop a comprehensive and contextual understanding of the current as well as the future direction of policy, law and practices in the relevant jurisdictions that will inform further work of the D2D CRC Research Program.

4. RESEARCH METHODS AND SOURCES OF DATA

4.1 Research Design

The project employed a mixture of legal and empirical research methods drawing on the following sources of data:

1. **Interviews** with key stakeholders, technologists, and users in each country in relation to their understanding of the capabilities and uses of Big Data, their perception of issues and challenges in relation to Big Data, their perception of existing and proposed strategies, policies, laws and practices, and their recommended responses to perceived challenges. Interviews were conducted face-to-face where possible and via Skype or video-conferencing where feasible.

2. **Laws**, regulatory frameworks governing, and practices relating to the use of Big Data in each country.

3. **Documentary and other sources** that provide further information on the technology and legal/policy responses.

The research was carried out in four stages in relation to each country:

Stage 1: This stage involved various preparatory work including employment of research assistants; collection and analysis of legal and documentary materials; application for ethics approval; design (or adaptation) of semi-structured interview instruments (see Technical Reference 1) in consultation with government agencies and other CRC partners; and contact with law enforcement and national security agencies to organise interviews.

Stage 2: This stage involved continuing the analysis of legal and documentary materials; interview with stakeholders; and transcription of interviews.

Stage 3: This stage involved the completion of all analyses and a draft country report, which was circulated to relevant agencies and experts for comment.

Stage 4: Following feedback from agencies and experts, the country reports were revised and results of the project submitted to the CRC and published in academic outlets.

4.2 Research Process

4.2.1 Empirical Analysis

The project received human research ethics approval from both UNSW Australia (Approval number 14 168) and from Deakin University (Approval number 2014-295) in December 2014.

The research team then worked with various government agencies and the CRC's government and industry partners to jointly scope interviews, identify key persons to interview and arrange access to appropriate officials in each country. Prospective research participants covered a range of stakeholders including law enforcement and intelligence officials, policymakers, computer technologists, and representatives of citizen groups.

Invitations were sent to prospective research participants by email, fax or post. Attached to each invitation was a Participant Information Statement and Consent Form which explained the project aims and various safeguards for research participants, including voluntary participation, confidentiality, anonymity and freedom to withdraw consent (see Technical Reference 2).

The recruitment process was quite long and drawn out, as responses to the invitations were not immediately forthcoming. For the Australian component, the Secretary of the Attorney-General's Department assisted by sending letters to the relevant agency heads endorsing the research project and suggesting that they encourage their staff to participate in the project when invited. Email reminders were sent to invitees when no responses were received. These were sometimes accompanied by phone calls. Some declined to take part. Others were no longer followed up after several attempts.

A total of 63 research participants took part in the research project: 38 from Australia (interviewed from 25 March 2015 to 13 November 2015), 14 from the UK (interviewed from 24 February 2016 and 18 March 2016) and 11 from Canada (interviewed from 15 October 2015 to 26 February 2016).

In order to anonymise research participants, details of participants' organisation or team were removed from the interview notes and transcripts. In such cases, the details were replaced by square brackets and an organisation type. For example, 'Australian Federal Police' may be replaced by '[law enforcement agency]'. In order to increase the fluency and relevance of selected quotations in the report, we have also used ellipsis and square brackets to indicate the omission or replacement of words respectively.

Where consent was given, interviews were recorded and a confidential transcription service was employed for the transcription of interviews. Notes were taken by researchers where research participants did not wish their interviews to be audio recorded. Where notes were taken, we attempted to mirror as closely as possible the words used by research participants in the quotes used in the report.

Although interviewers used a standard set of questions, there was some variation between interviews, which were semi-structured to allow a natural conversation between researchers and participants. On some occasions, research participants responded to earlier questions in broad terms that pre-empted later questions. In such situations, later questions were either skipped or repeated only to ask whether the research participant had anything further they wished to add in response. Accordingly, when interviews were coded, answers were sometimes collated from responses to different questions.

We classified research participants in accordance with their role and the nature of the organisation for which they worked. In each case, there were three potential classifications: Operational (O), Technical (T) and Policy (P). We listed these as 'Role/Organisation Type' so that, for example, T/O indicated a person with a technical role inside an operational organisation. We also introduced the classification of O-P/O, to capture those with senior roles in operational organisations whose role encompassed both management of operational personnel and strategic policy aspects. The Policy classification was broad, and included individuals and agencies with a legal or policy role, community organisations and NGOs and individuals and agencies with an oversight role over operational agencies. Where a research participant was being interviewed in relation to a recent former role, the coding matched the former role and organisation rather than current role and organisation.

We had three sets of overlapping questions that we used in the interviews. We labelled these Operational, Technical and Policy. The questions that research participants were asked generally aligned with their role and organisation. Some research participants were comfortable answering more than one set of questions. For example, a research participant classified as T/O may be comfortable answering both the Technical questions and the Operational questions. Some research participants preferred to only answer questions that aligned with their own role; for example one research participant classified P/T was only asked the Policy questions. Other research participants were able to answer some but not all

of the questions in relation to their organisation where this did not align with their own role. For example, one research participant classified T/O was able to answer the Technical questions and some of the Operational questions. In one case, a research participant classified T/O was also willing to answer Policy questions. Where a research participant answered more than one set of questions, they were not re-asked questions that were identical or similar to questions they had already been asked. One issue with this approach (identified after the conclusion of the Australian interviews) was that the T/O group were sometimes asked slightly different formulations of the same question, depending on which questions they answered first. This was rectified by standardising the form of the question for the Canadian and UK interviews.

Some research participants insisted on group interviews with more than one person answering questions in a single interview. Where research participants in group interviews shared the same classification (e.g. T/O), we used the appropriate set of questions. Where research participants had different roles, the questions asked were adjusted to incorporate as much as possible within the time available.

In addition to the classification of research participants based on role and organisation, we also introduced other classifications that were relevant to our analysis for some sections of the report. Operational organisations were classified into those focussing primarily on intelligence gathering ('**INTEL**') and those focussing primarily on law enforcement ('**LE**'). Further, organisations were also classified by sector into three groups – Research / Private / NGO, Government (including state and federal government), and Independent (for formally independent but government-established offices and agencies; for example an independent commissioner or oversight body). Where a research participant was being interviewed in relation to a recent former role, classification was based on this former role.

Table 4-1 summarises the numbers of research participants by type of organisation for each country. All were selected because they were able to provide relevant information on the use of data or regulation of data use for law enforcement or security intelligence. Time and resource constraints necessitated relatively modest sample sizes for the UK and Canada components of the study. Although the Australian, UK and Canadian samples were drawn from broadly similar types of organisations (technical organisations were excluded in the UK and Canada), there were some notable differences in the research participants. The three samples may therefore not be directly comparable.

Table 4-1: Number of Research Participants by Organisation Type and Country

	Operational	Technical	Policy	Total
Australia	19	7	12	38
UK	5	-	9	14
Canada	6	-	5	11
Total	30	7	26	63

It is important to emphasise that the empirical findings presented in this section provide a snapshot of the *views and perceptions* of research participants only. These views and perceptions may or may not be based on a comprehensive or accurate understanding of the issues involved. The findings are therefore meant to be *indicative*, rather than representative, of the views of the populations of stakeholders in these countries.

In the empirical chapters of the country reports, conclusions drawn from interviews are provided in 'Summary and Implications' boxes.

The goal of our empirical research is to capture understandings, perceptions and views of individual research participants on a range of issues. I**t is important to emphasise that the empirical findings presented in each report provide a snapshot of the *views and perceptions* of research participants only. These views and perceptions may or may not be based on a comprehensive or accurate understanding of the issues involved. Given that the sample size is relatively small and not necessarily representative of the population of stakeholders, the findings are meant to indicate issues and not to be read as a comprehensive coverage of all relevant information. We do not attempt in our research to evaluate or correct research participants' views, although we include cross-references to other sections in each report where appropriate.**

4.2.2 Legal and Policy Analysis

The legal and policy analysis was focused on the identification of features of the legal framework in each of jurisdiction that were particularly relevant to the use of advanced analytics and large data sets for law enforcement and national security purposes in that jurisdiction. This analysis reflects the law as at 31 March 2016 (Australia), 26 May 2016 (UK) and 1 June 2016 (Canada).

To assist in identifying those features the research team developed a broad set of indicators of a legal framework that would enable or require desirable and effective analytical practices. The development of the set of indicators and their application is discussed in Chapter 5.

The scope of the study also required the researchers to identify those legal instruments in each country that are particularly relevant to the research questions of this study. The selection of the relevant set of instruments is discussed in context in each country study.

In the UK and Canada Reports, conclusions drawn from the legal analysis are provided in 'Summary and Implications' boxes. In the Australia Report, relevant insights from the legal analysis are provided in 'Observations' boxes.

5. INDICATORS OF A LEGAL AND POLICY FRAMEWORK THAT SUPPORTS 'DESIRABLE AND EFFECTIVE' BIG DATA PRACTICES

The policy and legal reviews conducted in this study considered the features of the legal landscape and in particular elements of the legal framework that either supported or potentially prevented desirable and effective practices in the use of advanced analytics and large data sets for law enforcement and national security purposes. What 'desirable and effective Big Data practices' may entail for Australian national security agencies has not been established by the Australian government. The research team therefore identified a number of provisional high level indicators. According to the technology as well as the legal and policy researchers of the D2D CRC, these would point to the existence of such a framework should all of the indicators be present.

These indicators provided a lens through which the current legal and policy framework could be viewed and assessed. The indicators were, however, identified as provisional as it was acknowledged by all stakeholders that their identification and their application in this study will stimulate debate, enrich the analytical framework for the discussion of these elements and lead to a refinement of the concept of 'desirable and effective' practices in this context. The indicators were not viewed as complete or comprehensive. The research team therefore retains an open mind as to what might be a more complete set.

5.1 Development of the lens

The legal and policy analysis for this study was undertaken to answer the following research questions:

- What are the strategies, policies, laws, regulatory frameworks, practices and technologies relevant to Big Data adopted by Australia/the UK and Canada, and are they perceived as effective in meeting social, legal or policy objectives, including, but not limited to, law enforcement and national security objectives?
- What responses are needed to deal with potential challenges?

Early in the study, it became clear that Australia has not adopted Big Data-specific strategies, policies, laws, regulatory frameworks, practices and technologies relating to law enforcement and national security. However, it does have a host of measures that are directly or indirectly relevant to Big Data application. To assist in differentiating between what was of greater or lesser relevance to Big Data in this context the team developed a mechanism that reflects not only what would be indicative of effective practices but also what would be desirable to balance the different policy objectives that are relevant in this context.

The initial set of indicators, though not necessarily complete, is indicative of an appropriate overarching legal and policy framework. Collectively the presence of these indicators would indicate a framework that can support the effective use of advanced analytics and large data sets for law enforcement and national security purposes, while respecting the rights and interests of all stakeholders (including data subjects, the broader community and the economy), addresses proportionality and evidence-based justification, and ensures comprehensive identification and management of risk and opportunities.

Such a framework would in general enable sufficient, controlled access to and analysis of data for identified purposes, ensuring good governance while guarding against program management and operational risks such as overreach, intrusive practices with little or no real benefit, or 'scope creep' (continual incremental extension to new uses).

After initial formulation by the researchers, the D2DCRC management (combining technology and national security expertise) joined in shaping and refining the indicators. The

development was done on the understanding that the indicators are provisional and that further development and refinement may be expected during the course, and as a result, of the study.

The following indicators were identified:

5.1.1 Is access for data mining enabled?

The framework would enable access to relevant datasets held by government agencies (domestically and internationally), to open source data and to relevant privately-held data in a manner that allows data mining subject to the governance and control mechanisms set out 2 – 7 below.

The enquiry into access to data also extends to the use and abuse of mechanisms such as encryption, for instance key escrow and end-to-end encryption.

5.1.2 Are legal controls comprehensive and proportional?

The framework would ensure that proportional legal controls inform and guide the design, operation and management of (a) data mining and analysis, (b) data collection, (c) data retention and deletion, (d) data aggregation and (e) disclosure (domestically and, where required, internationally), weighing law enforcement and national security objectives as well civil liberties, other legal rights and individual and commercial interests.[112]

In the context of law enforcement and national security agencies, the test of proportionality (further discussed below), requires the decision-maker to assess whether the proposed action or measure fits within the statutory purpose, and then balance actual benefits, necessity or opportunity on the one hand, and costs and risks on the other, including risks projected onto others. This balancing process should be robust, qualitative and quantitative; it should lead to measures suited to achieve an identified and appropriate objective, that are necessary for achieving that objective (taking into account feasible alternatives or variations), and that limit imposition of disproportionate burdens on affected individuals[113] and society compared to the benefits actually achieved.

5.1.3 Are legal rules clear, principle-based, consistent and instructive?

The legal rules of the framework would be clear, principle-based and consistent, providing officials with appropriate guidance to take reasonable decisions and perform their functions correctly and efficiently in a dynamic environment. These rules would be challenged to provide 'future proofing' while being specific enough to avoid ambiguity, and so maintain auditability. This also implies a consistent employment of the principles in relation to different agencies and data sets where objectives and risk are sufficiently similar.

5.1.4 Is integrity of data and analysis supported?

The framework would support the integrity of data collected, retained and accessed by government for law enforcement and national security purposes, and the integrity of analytical and decision-making uses of such data and systems. Where integrity is assessed as low, data usage will reflect that fact and the principles in 5.1.6 will apply.

---

[112] While the researchers agreed that it was important to consider 'other legal rights and individual and commercial interests' the vagueness of this phrase was appreciated. It was identified as one that would need to be revisited once the study enabled it to be defined with greater clarity.

[113] It was agreed to consider any differences in protection enjoyed by Australian citizens and residents and foreign citizens.

5.1.5 Are data and systems protected?

The framework would protect the security of relevant data and systems. Appropriate measures would enable individuals to report concerns or breaches internally in a manner that supports integrity and governance.

5.1.6 Is accountability maintained?

The framework would ensure that access to data and data analysis and use for decisions is tracked and audited for justification, security and intrusiveness, and that decisions are subject to appropriate internal governance as well as independent oversight and accountability.[114] Decision-makers should remain accountable for their decisions, which imply an understanding of the provenance and integrity of data and awareness of any biases in the analytic process.

5.1.7 Are principles and rules regularly reviewed?

The framework would require the regular, transparent review of principles and rules to ensure that the system delivers intended results efficiently and reliably, proportional to impacts on civil liberties, other legal rights and individual and commercial interests.

5.1.8 Is there a sufficient measure of transparency?

To the extent consistent with the need for operational secrecy, the framework would ensure that the nature of data accessed, analytic processes employed, and who has access are as transparent as feasible for those potentially affected by decisions, and those with an interest in policy- and rule-making.

[114] 'Appropriate' in this document means reasonable and justifiable in an open and democratic society. Where relevant, it reflects proportionality principles.

TECHNICAL REFERENCES

Participant Information Statement and Consent Form

The following 4 pages contain the Participant Information Statement and Consent Form provide to the participants.

UNSW HREA Approval No: 14 168
Deakin HREC Approval No: 2014-295

THE UNIVERSITY OF NEW SOUTH WALES, DEAKIN UNIVERISTY AND DATA TO DECISIONS COOPERATIVE RESEARCH CENTRE (CRC)

PARTICIPANT INFORMATION STATEMENT AND CONSENT FORM

Big Data Technology and National Security:
Comparative International Perspectives on Strategy, Policy and Law

Janet Chan, Louis de Koker et al.

Introduction
You are invited to take part in this research project, which is called *Big Data Technology and National Security*. The study is being conducted by the following researchers: Professor Janet Chan, UNSW; Professor Louis de Koker, Deakin University; Professor Danuta Mendelson, Deakin University; Associate Professor Lyria Bennett Moses, UNSW; Dr Alana Maurushat, UNSW; Mr David Vaile, Deakin University; Mr Mike Gaffney, Attorney-General's Department; Mr Gregory Sadler, Attorney-General's Department; and Mr Patrick Grierson, Attorney-General's Department. You have been invited because you have knowledge of the technical, legal, policy or social issues relating to the use of Big Data for law enforcement and national security.

This Participant Information Sheet/Consent Form tells you about the research project. It explains the processes involved with taking part. Knowing what is involved will help you decide if you want to take part in the research.

What is the purpose of this research?
The main aim of this project is to examine the policies, regulatory approaches, processes and strategies used by Australia, the UK and Canada to balance the management and exploitation of Big Data for law enforcement and national security purposes, while safeguarding confidentiality and security of sensitive personal information, as well as the accuracy of data sets. The findings will be examined for their relevance and applicability to Australia. The focus will be on the three countries (above) as members of the Five Eyes intelligence community. In particular, we would like to canvass the views of key stakeholders, technologists and users in each country regarding the use of Big Data for law enforcement and national security. This research project is not concerned with information that may be private, confidential, classified, or in relation to specific criminal offences. You are requested not to disclose such information during the interview.

Description of study procedures and risks
If you decide to participate, you will need to sign and return the consent form to the research team at <d.cater@unsw.edu.au>. One of the researchers will contact you to organise an interview in person, by telephone or video, at a time and a place convenient to you and the researcher. The interview will take approximately one hour. With your consent, notes will be taken, the interview will be digitally recorded and a confidential transcript of the interview will be prepared. If you consent to recording of the interview, you may also request to review a copy of the transcript.

What are the possible benefits of taking part?
The project will inform Australia's policy and approaches to the use of Big Data for national security and law enforcement. We cannot and do not guarantee or promise that you will receive any benefits from this study.

What are the alternatives to participation?
Participation in this research is voluntary. If you don't wish to take part, you don't have to. Your decision not to participate will not affect your future relations with the University of New South Wales, Deakin University or the Data to Decisions Cooperative Research Centre.

Confidentiality and disclosure of information
Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission, except as required by law. If you give us your permission by signing this document, we plan to publish the results as technical reports for the CRC and journal papers for the academic community. In any publication, information will be provided in such a way that you cannot be identified.

Complaints
Complaints may be directed to the Ethics Secretariat, The University of New South Wales, SYDNEY 2052 AUSTRALIA (phone (02) 9385 4234, fax (02) 9385 6222, email humanethics@unsw.edu.au). Any complaint you make will be investigated promptly and you will be informed out the outcome.

Feedback to participants
If you would like to receive a summary of the research findings at the completion of this study, please tick the box below on the Consent Form.

Your consent
Your decision whether or not to participate will not prejudice your future relations with the University of New South Wales, Deakin University or the Data to Decision CRC. If you decide to participate, you are free to withdraw your consent and to discontinue participation at any time without prejudice.

If you have any questions, please feel free to ask us. If you have any additional questions later, Professor Janet Chan (telephone 0401 713 461 or email J.Chan@unsw.edu.au) will be happy to answer them.

You will be given a copy of this form to keep.

PARTICIPANT INFORMATION STATEMENT AND CONSENT FORM (continued)

Big Data Technology and National Security:
Comparative International Perspectives on Strategy, Policy and Law

Janet Chan, Louis de Koker et al.

Declaration by Participant

I have read the Participant Information Sheet or someone has read it to me in a language that I understand.

I understand the purposes, procedures and risks of the research described in the project.

I have had an opportunity to ask questions and I am satisfied with the answers I have received.

I freely agree to participate in this research project as described and understand that I am free to withdraw at any time during the project without affecting my future care.

I understand that I will be given a signed copy of this document to keep.

……	……
Signature of Research Participant	Signature of Witness
……	……
(Please PRINT name)	(Please PRINT name)
……	……
Date	Nature of Witness

☐ I consent to audio recording of the interview

☐ I consent to note taking during the interview

☐ I would like to review a copy of the interview transcript (if consent to audio recording)

☐ I would like to receive a summary of the research findings at the completion of the study. (Please provide an email address below).

REVOCATION OF CONSENT

Big Data Technology and National Security: Comparative International Perspectives on Strategy, Policy and Law

Janet Chan, Louis de Koker et al.

I hereby wish to **WITHDRAW** my consent to participate in the research proposal described above and understand that such withdrawal **WILL NOT** jeopardise any treatment or my relationship with The University of New South Wales, Deakin University or the Data to Decision CRC.

…… …….

Signature Date

……

Please PRINT Name

The section for Revocation of Consent should be forwarded to:

Professor Janet Chan
Law School, Building F8, UNSW Australia, Sydney NSW 2052, Australia
j.chan@unsw.edu.au

Interview instrument – Australia

Big Data Technology and National Security

A. Interviews with law enforcement and intelligence officials – AUSTRALIA

Introductory Information

Thank you for agreeing to an interview. Before we start I would like to tell you a bit more about the study and what we hope to achieve from the interviews. As you would be aware from the Participant Information Statement, the main aim of this project is to examine the policies, regulatory approaches, processes and strategies used by Australia to balance the management and exploitation of Big Data for law enforcement and national security purposes, while safeguarding confidentiality and security of sensitive personal information, as well as the accuracy of data sets. This research project is not concerned with information that may be private, classified, or in relation to specific criminal offences. You are requested not to disclose such information during the interview. We would like to also remind you that anything we talk about will remain confidential to the project and if we use anything you say in this interview in our publications, we will make sure that you will not be identifiable.

Current Position

1. Please describe the responsibilities of your current position and the organisation and team or unit in which you are employed.
   [NB: if you have any concerns about being identified, we will not include specifics about your position/responsibilities that could be used to identify you.]

2. Could you tell us (a) what your education and training background is, and (b) about your work experience prior to the current position?

General

3. When does digital/computer technology hinder you in your work and when is it particularly helpful?

Data Sources, Access and Sharing

I am going to ask some questions around data sources, access and sharing. In these questions, I use the term 'data' broadly to capture records, information and intelligence.

4. What types of data do you (or your unit) use in your work?

5. What types of data do you (or your unit) generate in your work? How is this captured?

6. Does your unit share data with other agencies, and if so, which ones?

7. What are you major concerns in relation to data access from other agencies or sharing data with other agencies?

8. Do these problems affect your (or your staff's) morale or sense of professionalism?

9. How can these problems be overcome?

Data Analytics

10. What do you (or your unit) mainly use these data for?

11. How do you (or your unit) normally use data for [law enforcement investigation/crime prevention/security intelligence etc] as described above?

12. Do you (or your unit) do data visualisation or data analysis? If so, what techniques or software do you(or your unit) use? Are these off-the-shelf or custom tools?

13. What are the most serious issues/problems that may prevent you (or your unit) making greater use of data analytics?

14. Do you think your agencies' access to data and analytical tools is better or worse that other agencies, your foreign counterparts and the private sector?

Big Data

15. This research project is concerned with the use of Big Data. The term 'Big Data' has been used by people in a number of ways. How would you define Big Data?

16. As far as you know, what is Big Data capable of doing that 'ordinary data' can't?

17. To what extent are you (or your unit) making use of Big Data tools in your work?

18. What are the most serious issues/problems that may prevent you (or your unit) from making more use of Big Data?

19. What do you see are the risks of using Big Data for law enforcement or security intelligence?

Regulation

20. As you know, there are laws, regulations or procedures governing the use of data by law enforcement or security agencies. In your view are these laws, regulations, procedures, guidelines etc appropriate? Are they effective?

21. How do you think the law should strike a balance between privacy/individual rights and public concerns such as national security, terrorism and serious crimes?

Thank you very much for taking the time to be interviewed. Your input is much appreciated.

Big Data Technology and National Security

B. Interviews with technologists/designers – AUSTRALIA

Introductory Information

Thank you for agreeing to an interview. Before we start I would like to tell you a bit more about the study and what we hope to achieve from the interviews. As you would be aware from the Participant Information Statement, the main aim of this project is to examine the policies, regulatory approaches, processes and strategies used by Australia to balance the management and exploitation of Big Data for law enforcement and national security purposes, while safeguarding confidentiality and security of sensitive personal information, as well as the accuracy of data sets. This research project is not concerned with information that may be private, classified, or in relation to specific criminal offences. You are requested not to disclose such information during the interview. We would like to also remind you that anything we talk about will remain confidential to the project and if we use anything you say in this interview in our publications, we will make sure that you will not be identifiable.

Current Position

1. Please describe the responsibilities of your current position and the organisation in which you are employed.
   [NB: if you have any concerns about being identified, we will not include specifics about your position/responsibilities that could be used to identify you.]

2. Could you tell us (a) what your education and training background is and (b) your work experience prior to the current position?

Big Data – Capabilities

3. This project is concerned with the use of Big Data. The term 'Big Data' has been used by people in a number of ways. How would you define Big Data? What does it mean in the context of your work? How does it relate to other terms you might use?

4. Do you use Big Data in your work/systems? What role do Big Data techniques have in your work/systems? What is your role in relation to these?

5. What types of data analysis do you or does your organisation do? What are the outputs? How reliable/accurate are these outputs?

6. Who are the users of your product/system? Who can access data within the system? What mechanisms are used to ensure the security and privacy of data within the system?

7. In your organisation/system, what are the main challenges you face with respect to Big Data or data analysis?

8. What do you see are the opportunities or possibilities that [data analysis/data science if they used these terms] or Big Data can open up for law enforcement and security intelligence?

9. What data analytic and data visualization tools or software do you use when dealing with large data sets? What do these tools provide you with? Are they useful?

10. What do you know about Big Data that everyone in the field will know in five years? How would you improve the way your organisation designs systems or builds tools for working with national security and law enforcement data?

11. What issues do you think are likely to come up in the future for Big Data or data analytics? What advice would you give policy-makers on the use of Big Data or data analytics/data science for law enforcement or national security purposes?

Risks

12. What are the risks of using data analytics to support law enforcement and enhance national security relating to your work or product? Who is exposed to these risks? How should these risks be managed?

13. How important is it that outputs of your system are reliable and accurate? How tolerant of invalid, unreliable, corrupted or non-relevant data can your system afford to be?

Design Issues

14. Who sets the design parameters for your product/system?

15. To what extent can some of the risks of data analytics be mitigated through the design of analytical tools? For each of the following issues, please indicate whether you take it into account in your design (Yes/No), and if so, how:

Issues	Yes/No	How is it taken into account?
Protection of privacy		
Communications confidentiality		
Personal information security		
Data integrity, [ie that information held is accurate, current, relevant and not misleading]		
Regulatory compliance		
Testing and evaluation		
Comprehensibility to decision-makers		
Avoiding unintended consequences		
Avoiding discrimination		
Potential for de-identified data to be re-identified		
Agency inter-operability		
Cost		

Thank you very much for taking the time to be interviewed. Your input is much appreciated.

Big Data Technology and National Security

C. Interviews with policymakers, citizen groups – AUSTRALIA

Introductory Information

Thank you for agreeing to an interview. Before we start I would like to tell you a bit more about the study and what we hope to achieve from the interviews. As you would be aware from the Participant Information Statement, the main aim of this project is to examine the policies, regulatory approaches, processes and strategies used by Australia to balance the management and exploitation of Big Data for law enforcement and national security purposes, while safeguarding confidentiality and security of sensitive personal information, as well as the accuracy of data sets. This research project is not concerned with information that may be private, classified, or in relation to specific criminal offences. You are requested not to disclose such information during the interview. We would like to also remind you that anything we talk about will remain confidential to the project and if we use anything you say in this interview in our publications, we will make sure that you will not be identifiable.

Current Position

1. Please describe the responsibilities of your current position and the organisation in which you are employed.
   [NB: if you have any concerns about being identified, we will not include specifics about your position/responsibilities that could be used to identify you.]

2. Could you tell us (a) what your education and training background is and (b) your work experience prior to the current position?

Big Data – Capabilities

3. This project is concerned with the use of Big Data. The term 'Big Data' has been used by people in a number of ways. How would you define Big Data?

4. As far as you know, what is Big Data capable of doing that ordinary data can't?

5. To what extent is Big Data currently being used for law enforcement and security intelligence in Australia?

6. Do you think Australian agencies access to data and analytical tools is better or worse that their foreign counterparts and the private sector?

7. In your view should this use by Australian agencies be expanded? If so, in what way should this be expanded and what could be achieved? If not, why not, and what would be the implications?

8. What do you see are the opportunities or possibilities that Big Data can open up for law enforcement and security intelligence? How can these possibilities be delivered? What are the barriers and how could they be overcome?

Big Data – Regulation

9. What are the challenges and risks of Big Data technology to support law enforcement and enhance national security? Who is exposed to these risks? How should these challenges and risks be managed?

10. How is the use of Big Data currently being regulated? What laws, policy, codes of practice, standards, etc. are in place in this jurisdiction? How effective are they? What are their shortcomings?

11. [If appropriate] Do you have a sense of the history of these regulations? If yes, why are they the way they are?

12. What accountability, transparency or oversight mechanisms are in place? Are they appropriate and effective?

13. What protections, if any, should remain in place in circumstances where an individual consents to the use or sharing of their data?

14. What future strategies are required for Big Data?

Scenario:

15. Now I would like to ask you a series of questions in relation to a hypothetical scenario:

Lucy is an 8 year old girl who has been kidnapped from her home in Lane Cove in Sydney. All avenues of traditional physical surveillance and canvassing of the area so far haven't produced any leads.

How do you feel about the immediate and expeditious use of big data tools in these circumstances?

What difference would it make if this was not just a kidnapping but there's suspicion of paedophilia? What difference would it make if there was suspicion that the kidnapping was linked to terrorism, with an intent to blow Lucy up in a public place?

16. To what extent should considerations such as privacy give way in the face of serious, imminent threats such as child kidnapping, child sexual abuse or terrorism?

17. Assuming that Big Data had been proven effective in other instances, are there ways that Big Data techniques could be used appropriately? What kind of laws, regulations, accountability mechanisms would need to be in place?

18. To what extent should there be transparency about the nature of data collected or the algorithms employed in analysis, both within an agency or more broadly?

19. How do you think your / your organisation's views about the design and regulation of Big Data technology align with the views of other stakeholders? Do you have any thoughts on how any conflict might be resolved?

20. To what extent are your views shaped by internal or personal experience as opposed to external sources such as blogs, watch groups and media?

Thank you very much for taking the time to be interviewed. Your input is much appreciated.