FUNDING AUSTRALIA’S FUTURE

BIG AND BETTER DATA, INNOVATION AND THE FINANCIAL SECTOR

DR IAN OPPERMANN

OCTOBER 2015

fundingaustraliasfuture.com
Funding Australia’s Future

The Australian Centre for Financial Studies (ACFS) instigated the Funding Australia’s Future project in 2012 to undertake a stocktake of the Australian financial system and analyse its role in facilitating economic growth within the wider economy.

In an economy which has enjoyed 24 years of consecutive economic growth and shown a resilience through the Global Financial Crisis (GFC) which was the envy of many nations, the financial sector has played a strong and pivotal role. The past decade, however, has been one of significant change. The growth of the superannuation sector, the impact of the GFC, and the subsequent wave of global re-regulation have had a profound effect on patterns of financing, financial sector structure, and attitudes towards financial sector regulation. Identifying the extent to which these changes are transitory or likely to be more permanent is crucial to understanding how financing patterns and the financial sector will develop over the next decade or so.

Stage Three of Funding Australia’s Future explores three specific challenges to the financial sector highlighted by the Financial System Inquiry, Tax System Review and Intergenerational Report. While diverse, each of these topics has a bearing on the future of the financial system and its role serving the economy.

In undertaking this analysis, ACFS has worked with a group of financial sector stakeholders comprising Accenture, the Association of Superannuation Funds of Australia, Challenger Limited, IBM, Industry Super Australia, National Australia Bank, Self Managed Super Fund Association and Vanguard Investments, as well as the Treasury.

This paper is one of three in Stage Three, which includes:

1. Dividend imputation and the Australian financial system: What have been the consequences?
   *Professor Kevin Davis, University of Melbourne and Australian Centre for Financial Studies*

2. Big and better data, innovation and the financial sector
   *Dr Ian Oppermann, CSIRO*

3. Financial issues in retirement
   *Professor Deborah Ralston, Monash University*

All Funding Australia’s Future papers can be accessed through the Funding Australia’s Future website: [www.fundingaustraliasfuture.com](http://www.fundingaustraliasfuture.com)
**Notes on the authors**

**Professor Kevin Davis:** Kevin Davis is Professor of Finance at the University of Melbourne, Research Director of the Australian Centre for Financial Studies and a Professor of Finance at Monash University. His primary research interests are financial regulation, financial institutions and markets, financial innovation and corporate finance. He is co-author/editor of 16 books in the areas of finance, banking, monetary economics and macroeconomics and has published numerous journal articles and chapters in books. He is the Deputy Chair of SIRCA, a member of the Australian Competition Tribunal, and has undertaken an extensive range of consulting assignments for financial institutions, business and government. Professor Davis is a Senior Fellow of Finsia, a Fellow of FTA and holds Bachelor of Economics (Hons I) from Flinders University of South Australia and a Master of Economics from the Australian National University. He was appointed by the Federal Treasurer in December 2013 as a panel member of the Financial System Inquiry chaired by Mr David Murray.

**Dr Ian Oppermann:** Ian Oppermann has over 20 years experience in the ICT sector and, for 10 years, has led organizations with 100 people or more, delivering products and outcomes that have impacted hundreds of millions of people globally. He has held senior management roles in Europe and Australia as Director for Radio Access Performance at Nokia, Global Head of Sales Partnering (OSS) at Nokia Siemens Networks, and then Divisional Chief and then Flagship Director at CSIRO and CEO of Rozetta Technology, a knowledge infrastructure ‘Big Data’ company.

Ian is considered a thought leader in the area of the Digital Economy and is a regular speaker on ‘Big Data’, broadband enabled services and the impact of technology on society. He has contributed to six books and co-authored more than 120 papers. Ian has an MBA from the University of London and a Doctor of Philosophy in Mobile Telecommunications from Sydney University. Ian is a Fellow of the Institute of Engineers Australia, a Fellow of the IEEE, a Fellow of the Australian Academy of Technological Sciences and Engineering, a Senior Member of the Australian Computer Society, and a member of the Australian Institute of Company Directors.

**Professor Deborah Ralston:** Deborah Ralston is Professor of Finance at Monash University and Cluster Lead CSIRO-Monash Superannuation Research Cluster. Previous appointments include Pro Vice Chancellor of the Division of Business Law and Information Science at the University of Canberra, and Director of the Centre for Australian Financial Institutions at the University of Southern Queensland. Deborah’s academic background is in economics and financial management and she has taught in the areas of economics, corporate finance, credit risk management, financial statement analysis and financial institutions management. Her research interests include the impact of financial regulation, the strategy and management of financial institutions and regional economic development. She has published widely in these areas and is a co-author of the text Financial Institutions Management.

Deborah is a Fellow of the Australian Institute of Company Directors, the Financial Services Institute of Australasia, and the CPA Australia. Deborah was formerly a Director of Heritage Building Society 1992-2003 and is currently a Director of the listed mortgage broking company Mortgage Choice.
Contents

Funding Australia’s Future.................................................................................................................... 2
Notes on the authors ................................................................................................................................. 3
Executive Summary................................................................................................................................. 6
1. Introduction – it’s about productivity................................................................................................. 7
2. Australia’s productivity challenge – the link to bigger, better data ............................................. 8
   2.1 Reframing productivity for the digital economy ........................................................................... 11
   2.2 Bigger, better data as a driver of productivity ............................................................................ 14
   2.3 Accounting for value of data - the need for a framework............................................................. 15
3. Some considerations when applying data and analysis ................................................................. 16
   3.1 Simplify at your own risk .............................................................................................................. 16
   3.2 Unexpected correlations, human bias and noise ......................................................................... 17
   3.3 The strength of distant data ......................................................................................................... 19
   3.4 Variety – how wide to cast the net? ............................................................................................... 20
4. Coming to grips with the digital economy – moving bits not atoms ........................................... 22
   4.1 Digitise ....................................................................................................................................... 23
   4.2 Link ............................................................................................................................................ 24
   4.3 Model (and analyse and optimise) ............................................................................................... 25
   4.3 Personalise .................................................................................................................................. 26
   4.4 The changing nature of work ..................................................................................................... 27
5. The changing financial sector .......................................................................................................... 27
   5.1 Levelling information asymmetry – innovate or get disrupted..................................................... 29
   5.2 Know your customer, know your product (then join the dots)................................................... 29
   5.3 Peer-to-peer lending – a new business model ............................................................................ 30
   5.4 Crowd funding – extending the peer-to-peer concept ................................................................. 31
   5.5 Robo-advice – automating and personalising analysis of financial data .................................... 33
   5.6 Feeding the robo-traders – big data for high frequency trading .............................................. 34
   5.7 Investing for productivity growth – an unmet need for infrastructure data ............................. 35
   5.8 Assuring integrity by detecting anomalies .................................................................................. 37
   5.9 Personal data as a raw material - profiting from sharing............................................................ 37
6. Digital transformation in government .............................................................................................. 38
   6.1 Transforming government through data....................................................................................... 38
   6.2 Open government data ................................................................................................................ 39
6.3 Sharing data within government ................................................................. 40
6.4 Broader privacy considerations in an Australian context.......................... 42

7. Conclusion – Australia has a productivity challenge .................................. 43
  7.1 Linking ‘bigger and better data’ to the Financial System Inquiry’s recommendations ... 44

References ........................................................................................................... 48

About the Australian Centre for Financial Studies ........................................... 53

ACFS Research Code of Conduct ..................................................................... 53
Executive Summary

Australia has a predominantly services driven economy which is increasingly digital, online and driven by user generated data. Within our digital economy, financial services – principally banking, insurance and capital markets – has overtaken all other sectors to become the largest sector of the economy.

The recent financial crisis and its aftermath have fundamentally challenged beliefs about the way financial systems, businesses and households behave. There are increasing demands for financial services organisations to have a greater customer-centric focus, operate with greater transparency and provide greater consumer protection. Customers have also demonstrated a far greater willingness to engage with non-traditional loan, investment, saving and retirement products. As a consequence, there have been waves of technology driven innovation in the finance sector coming from both incumbents and from new entrants challenging for a share of the market.

In such an environment, data are significantly undervalued as a factor of production, a means of increasing transparency, and a source of innovation. The ability to harness a wide range of large, constantly evolving and highly personalised data sets is a strong driver of productivity and supports the creation of new, high value services. As financial services companies continue to create value by digitally enabling their current business models, the opportunity exists to adopt new, disruptive business models targeted at ever more refined customer segments.

This paper provides recommendations focused on the increased use of bigger, better data, providing examples of innovation and disruption from a number of sectors including finance. Amongst the recommendations, the most important for Australia include the urgent need for regulatory clarification – facilitating the use and benefits from bigger, better data while protecting consumer privacy – and the need to open up data sets for research and commercial purposes within such a clarified regulatory environment.

The author would like to thank the members of the Steering Committee for their contributions as well as Therese Catanzariti (Barrister), Asoc. Prof. Kimberlee Weatherall (Sydney University), Jim Kotoulas (Industrie IT) and Drew Clark (Dept. of Communications).
1. Introduction – it’s about productivity

Australia is historically viewed as having an economy dominated by agriculture and resources. Indeed, Australia’s resilient economic position in the face of the long lasting global financial crisis (GFC) has largely been the result of the strength of the commodities markets coupled with steadily increasing demand from the emerging Chinese and Indian super-economies.

The reality, however, is that for decades Australia has been transforming towards a services dominated economy and our services economy is increasingly digital, online and driven by data. In all sectors, services are increasingly created, delivered and consumed through digital means driven by the increasing adoption of online, mobile, digital technologies. Recent years have witnessed dramatic changes in the way we consume music, watch movies, book holidays or choose hotels. Whilst less obvious to the average consumer, there are dramatic changes underway in all industry sectors, and in government, driven by changing consumer expectation, reduced barriers to new entrants and an increasingly borderless world of information flow.

Whilst acknowledging that our economy has fared relatively well with the underpinning of a strong and sustained resources demand, productivity growth in Australia has actually fallen significantly over the last decade across most industry sectors. This leaves Australia in a vulnerable position for a post-mining boom economy. We are already beginning to feel the effects of the softening of growth rates of China and the corresponding reduction in demand for Australian commodities. The need to lift productivity in our major services sectors is now brought into much sharper relief.

Compounding the effects of the decline in productivity growth, Australia must also anticipate the needs of an ageing population which will place increased demands on health and human service delivery. With a dramatic change expected in the ratio of retirees to employees, and the expectation of health care costs rising much faster than the consumer price index as recently reconfirmed by the 2015 Intergenerational Report, the need to improve productivity again comes to the fore. Recognising and responding to these challenges in an innovative way is essential if Australia is to be assured of a strong economic and social future (Department of Treasury 2015).

Increasing productivity in the digital services sector has become the new benchmark for international performance. Within our digital services economy, the financial sector has overtaken all others to become the largest sector of the Australian economy. A 2014 report from the Financial Services Council highlighted that the Australian financial services sector employed more than 400,000 people and contributed approximately $18 billion annually in corporate tax (Financial Services Council and UBS 2014). Changing customer attitudes and technology innovation are however eroding the position of traditional financial services companies.

The long running GFC and its consequences are driving a change in relationship between financial systems, businesses and households. There are increasing demands for financial services organisations to adopt a greater customer-centric focus, operate with greater transparency and provide greater consumer protection. At the same time, customers have become less ‘loyal’ to traditional organisations and are demonstrating a far greater willingness to engage with non-traditional loan, investment, saving and retirement products. As a consequence, there have been waves of technology driven innovation in the finance sector coming from both incumbents and
from new entrants challenging for a share of the market. A recent KPMG global survey highlighted the insurance sector in particular as struggling to innovate, with half of companies surveyed already being disrupted by new, more nimble competitors (KPMG 2015).

As with other segments of the economy, the clear opportunity exists for financial services companies to create significant value by continuing to digitally enable their current business models – leading to higher productivity – or by adopting new disruptive business models targeted at ever more refined customer segments. Consequently, there have been waves of technology driven innovation in the finance sector both from incumbents and from new entrants challenging for some of all of the banking process.

Today, the democratisation of data and the accessibility of advanced analytics have meant that smaller players can generate insight and create value by accessing large and varied data sets, and asking questions that were once out of reach. Importantly, it is also enabling new enterprises to reveal and more efficiently address the information asymmetry that is endemic in so many industries – from finance to retail and even public-service availability. In turn, this creates a range of new market opportunities for disruptive players and incumbents seeking to refresh their business models.

The next sections explore the productivity challenge in more detail and provide a link to the use of bigger, better data.

2. Australia’s productivity challenge – the link to bigger, better data

The sustained resources demand from Australia’s major trading partners has masked the impact of the decline in relative productivity in many industry sectors and saw Australia fall significantly behind other Organisation for Economic Cooperation and Development (OECD) countries. Australian labour productivity relative to the United States (US) went from almost 92 per cent in 1998 to just over 84 per cent in 2010 (Eslake and Walsh 2011). Whilst there have been recent increases in productivity, the rate of growth is well below those of our major trading partners. Table 1 from the Productivity Commission’s 2015 PC Policy Update shows the long term trends for labour and capital productivity (PC 2015). In the period 1973-74 to 2013-14, multifactor productivity (MFP) increased by an average of 0.8 per cent per annum. Over the same period, labour productivity increased by an average of 2.3 per cent, of which about two thirds (1.6 per cent) was attributable to ‘capital deepening’¹ with the remainder to multifactor productivity (MFP) growth.

By contrast, China experienced an average 5.1 per cent annual productivity growth in the period from 1985 to 2007 (Hamshmere et al 2014). In the most recent review of global competitiveness by the World Economic Forum,² Australia was ranked 21st in the world behind the US (5th), the UK (10th) and New Zealand (18th) (WEF 2014).

Table 1: Summary productivity statistics by market sector

<table>
<thead>
<tr>
<th></th>
<th>Long Term</th>
<th>Last complete</th>
<th>Period since</th>
<th>Latest years</th>
</tr>
</thead>
</table>

¹ See notes for figure 1: Multifactor productivity in financial and insurance services and the market sector.
² The World Economic Forum defines competitiveness as ‘as the set of institutions, policies, and factors that determine the level of productivity of a country’ (World Economic Forum 2014).
<table>
<thead>
<tr>
<th>Growth Rate</th>
<th>cycle</th>
<th>last cycle</th>
<th>2011-12</th>
<th>2012-13</th>
<th>2013-14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output (GVA)</td>
<td>3.0</td>
<td>4.0</td>
<td>2.4</td>
<td>4.5</td>
<td>2.6</td>
</tr>
<tr>
<td>Total inputs</td>
<td>2.3</td>
<td>4.0</td>
<td>2.5</td>
<td>3.4</td>
<td>2.2</td>
</tr>
<tr>
<td>Labour</td>
<td>0.7</td>
<td>2.4</td>
<td>0.0</td>
<td>0.2</td>
<td>-1.1</td>
</tr>
<tr>
<td>Capital</td>
<td>4.4</td>
<td>5.9</td>
<td>5.3</td>
<td>6.8</td>
<td>5.9</td>
</tr>
<tr>
<td>MFP</td>
<td>0.8</td>
<td>0.0</td>
<td>-0.1</td>
<td>1.1</td>
<td>0.4</td>
</tr>
<tr>
<td>Capital</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>deepening</td>
<td>1.6</td>
<td>1.6</td>
<td>2.5</td>
<td>3.1</td>
<td>3.3</td>
</tr>
<tr>
<td>Labour</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>productivity</td>
<td>2.3</td>
<td>1.6</td>
<td>2.4</td>
<td>4.2</td>
<td>3.7</td>
</tr>
<tr>
<td>Capital</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>labour ratio</td>
<td>3.7</td>
<td>3.5</td>
<td>5.2</td>
<td>6.6</td>
<td>7.0</td>
</tr>
</tbody>
</table>

Notes: Capital deepening is the change in the ratio of capital to labour, weighted by the capital share of market sector income. Labour productivity growth equals the sum of the growths of MFP and capital deepening.

Inter-period comparisons should be made with caution – productivity statistics for years prior to 2013-14 differ from the estimates released by the ABS in previous years, as a result of revisions by the ABS to historical hours worked series. This flowed through to measured labour productivity and multifactor productivity estimates. More details can be found in (Estimates of Industry Multifactor Productivity, 2013-14, Cat. no. 5260.055.002, December 2014).

Source: PC 2015.

The financial services sector has historically been a rapid adopter of information and communications technologies (ICT) and productivity has benefited significantly as a result. Figure 1 shows that for more than two decades, MFP in financial and insurance services has increased ahead of average MFP growth in the market sector of the Australian economy at 2.3 per cent per year compared with 0.9 per cent per year over the period 1989-90 to 2013-14. Productivity growth for other market sector industries combined averaged a meagre 0.5 per cent per year and has been declining since 2003-04. This has offset the increase in financial and insurance services, leading productivity in the market sector as a whole to decline over the 2003-04 to 2013-14 period.

**Figure 1: Multifactor productivity in financial and insurance services and the market sector**
The financial and insurance services sector is now the largest sector of the Australian economy at 9.9 per cent of gross domestic product (GDP) as measured in 2013-14. Mining (9.8 per cent), construction (9.3 per cent) and manufacturing (7.5 per cent) have all experienced a decline in MFP. Figure shows the average productivity growth for industries in the market sector for the period 2007-08 - 2013-14 as well as the growth in 2013-14. Finance and Insurance have shown significant growth in the last period measured at 3.3 per cent, however this is based on a 2007-08 – 2013-14 average of a more modest 1.1 per cent. From a whole of economy perspective, this seems weak against the background of 2007-08 – 2013-14 average declines in mining (-5.8 per cent) and no growth in manufacturing (0.0 per cent).

Figure 2: Industry contributions to market sector – multifactor productivity growth (2007-08 to 2013-14 and 2013-14)
Data from the Productivity Commission show the finance industry to be consistently making the largest investment in ICT, compared with other major industry sectors. The Australian Information Industry Association (AiiA) estimates that more than 90 per cent of total research and development (R&D) spend in the financial services industry is in ICT (AiiA 2015). With the exception of the ICT industry itself, the financial services industry has consistently been the largest single industry sector investing in ICT R&D.

Despite this investment, new technology means the financial services industry faces major changes to its traditional value chains. There is an ever greater emphasis on personalisation of services and a need to address niche markets. Innovative responses to customer requirements will become a critical point of differentiation and a means of responding to the new wave of disruptive technologies.

Access to data is central to the next wave of productivity gains and as a means of responding to these challenges.

2.1 Reframing productivity for the digital economy

Since the time of Adam Smith, we have defined productivity as a measure of the physical output produced from the use of a given quantity of inputs with an emphasis on land, labour and capital as the primary factors of production.

Over time, the broad term labour has been refined to distinguish entrepreneurial effort which gives a far greater return than simple labour. Capital has been refined to allow for financial capital, fixed and working capital, intellectual capital, human capital and social capital. The impact of technology is often considered simply as a means of improving the quality of these factors of production, especially capital. Investments in intellectual property including software can be viewed as building up intellectual capital. Intellectual property assets and software are often considered capital goods.

Source: PC 2015.
as they are non-financial, durable goods (they do not quickly wear out) that are used in the production of goods or services.

It is relatively easy to see how a licence for word processing software can be viewed as a durable capital good. For an additional employee, there is the need for an additional licence, or to ensure you within the maximum licence limit. Similarly, it is easy to understand this view if a company develops software tools for internal use. It dramatically underemphasises the role of very large, highly scalable software platforms which become the environment within which services are created for and by others. Facebook, SalesForce and Minecraft are examples of massively scalable platforms which provide a wide range of services from social networking to entertainment.

For definitional purposes, data can easily be treated like software as intellectual capital and as a durable good. This also substantially undervalues the magnifying role data can play in increasing productivity. A productivity model suited to manufacture of goods is ill suited to assess the contribution of data to service creation.

When the outputs of production are digital services rather than physical products, and the major inputs are data and digital services, the traditional model of understanding and measuring productivity becomes stretched. Modern services are increasingly created, delivered and consumed through digital means. The delivery and consumption of services in digital format greatly expands the geographic reach of service providers, crossing state and even national boundaries and allowing massive levels of data aggregation. End consumers of the services have also become the major source of the data driving the services they consume. The near limitless reproducibility and reusability of data, the low cost of storage and transmission, coupled with a high degree of software automation, have changed the role of data to become one of the primary factors of production for digital services companies.

Box 1: Uber

The world’s largest ridesharing company Uber does not own or lease the taxis it uses. It does not employ a global workforce of drivers or maintain a call centre to coordinate passenger pickup requests. It does not invest in physical assets you would expect from a traditional taxi company. It does employ staff to develop software platforms which automatically match available drivers with passengers, and which facilitates fare collection and driver payment. Through its smartphone application, Uber collects and processes vast quantities of data generated by millions of drivers, passengers and mobile global position system (GPS) devices. This is processed in its software platforms to make the ridesharing system work.

Uber also collects and analyses a great deal of other data to improve system performance, blending and analysing operational data along with data on availability of WiFi connections near users, mobile camera and even other applications running on a user’s smartphones, to improve both driver and passenger experiences.

In the traditional model of productivity, Uber does use land, labour and capital in the form of staff,

---

3 For more details see (Mojang 2015).
office space, power, software, computers and data storage. Their most important input is arguably the data generated by the apps running on the smart phones of passengers and taxi drivers around the globe. Uber’s ability to scale its output is only weakly dependent on the level of what are considered the major factors of production. Without user and driver generated data, their business would not function.


The traditional view of productivity also states that if a firm is not using its inputs at the maximum possible efficiency, there is scope to lower costs and increase profitability through productivity improvements. This may come about through the use of better quality inputs, adoption of technological advances, changes in work practices, or better management. This view of productivity is built on the notion that resources are finite, relatively inelastic and technology can be used to incrementally improve performance. If however the major inputs to a digital services company are data, this model must be rethought. Data can be reused or ‘consumed’ many times, in many different ways, for many different reasons without degrading the source data set. Data can also be shared or on-sold for other uses without diminishing the value to the original user.

Box 2: Thomson Reuters

Thomson Reuters is the world’s largest historical market data provider offering fine grained market activity (microsecond transactions) in the form of un-manipulated recorded trade and quote messages (Thomson Reuters, 2015a). These data were originally a by-product of trading activity and not seen as being intrinsically valuable. Today, the data are collected from all exchanges around the globe, and delivered to academic researchers, hedge funds, algorithmic traders and regulators. A single day’s data can range from 100s of gigabytes to close to a terabyte. Thomson Reuters engages labour to normalise the data into a common format, labour to develop and ensure the smooth running of the software platform delivering the data, uses capital in the form of communications systems and capital in the form of data centres to move and store data, and more labour in the form of sales staff to attract and retain customers.

The high levels of software driven automation means Thomson Reuter’s ability to increase its output is weakly dependent on increased inputs of land, labour and capital. Their exact same data product can be consumed by 100 or 1000 customers with minor adjustments to staff levels, office space or computing resources.

As the use cases grow for the data, the exact same data sets can be used repeatedly without impairing the value of the original. The same data can be used for academic research, training of high frequency trading algorithms, market research or as the raw material for assessing market efficiency. With an appropriate licence agreement, a user of the market data can also on-sell the exact same data for a completely different purpose.

Source: Thomson Reuters, 2015b.

Data-driven platform providers offering digital services can dramatically increase production and so productivity with minimal increases of land and labour, and with little increase in capital. Airbnb is one of the world’s largest accommodation providers and yet does not own or lease the homes and hotels it uses (Airbnb 2015). eBay is one of the world’s largest marketplaces and does not own or
lease space for these markets. These example companies have grown enormous revenues (and market capitalisations) by processing massive amounts of user generated data and providing personalised services.

2.2 Bigger, better data as a driver of productivity

The traditional statement of productivity is that output can be increased by using more inputs, making better use of the current level of inputs, technological improvements and by exploiting economies of scale. Multifactor productivity is then the residual portion of output growth that cannot be explained by increases in inputs including labour and capital. Defined as the growth in labour productivity minus the effect on productivity of a change in the amount of capital for each worker, MFP growth is typically explained by factors including advances in technology, rising education standards and changes in the socio-economic environment. When applied to data digital services, this view dramatically understates the role of data.

In 2009, Telstra commissioned a significant piece of research from ACIL Tasman looking at ICT as a driver of productivity growth (ACIL Tasman 2009a). The report stated that ICT influences labour productivity growth in Australia through its effect on MFP growth through technological changes, capital investment in ICT and falling ICT prices. The conclusion reflects the anecdotal view that ICT confers informational, strategic, transactional and transformational benefits. The report also cautioned that ICT is an enabler, a necessary but not sufficient condition for productivity growth and transformational improvements.

Less research has been done to evaluate the impact of data on productivity. In 2009 ACIL Tasman examined the impact of spatial data on economic activity and productivity in Australia and New Zealand (ACIL Tasman 2008; ACIL Tasman 2009b). It estimated that in 2008, the use and re-use of spatial information added $1.2 billion in productivity-related benefits to the New Zealand economy. This value was the result of increasing adoption of modern spatial information technologies over the period 1995-2008, and was equivalent to slightly more than 0.6 per cent of GDP in 2008. The results were similar for the Australian study with a modest but measurable impact from use and reuse of spatial data. The report cautioned that National Accounts do not capture the full extent of the spatial information industry and that spatial information was increasingly being used in most sectors of the economy where it is having a direct impact on productivity.

The 2009 reports are likely to have significantly underestimated the value of spatial data to the economy, partly due to the challenge of measuring the uses of data and partly due to the limitations of the productivity model itself. Since this report was released the uses of spatial data have increased significantly as high definition satellite images become available, better analysis techniques are developed, and as more cadastral and infrastructure information is digitised and mapped to a common spatial coordinates framework.

Spatial data sets arguably have the greatest potential to impact productivity, and Australian spatial data sets are likely to benefit Australian productivity more than any other country. In practice, all data ‘happens’ somewhere and so the spatial component of data associated with services, vehicles, machines, buildings, infrastructure, livestock, weather and people brings added richness to any existing service or facilitates new service creation. Spatial data are now used for navigation,
transport and logistics optimisation, infrastructure planning and maintenance, managing land use, biosecurity and environment management.

The productivity impact of spatial data remains a challenge to quantify in many cases. Knowing more accurately where a pipe is buried or having a single view of a property boundary shared by local, state and federal government agencies are challenging to measure. It requires a myriad of use cases to be considered and the counterfactuals to be explored.

In the world of financial services, spatial data are used by companies including CoreLogic for property pricing, and by Morgij Analytics for insurance risk assessment (CoreLogic 2015; MARQ 2015). The same spatial data sets can be used by many people, for many different purposes creating new and different services. When spatial data sets are linked together, or are linked with other data such as weather information, traffic flows, demographic information or predicted jobs growth, many new services can be created without diminishing the value of the underlying data.

The widespread availability of spatial data held within a multitude of Australian state and federal data agencies would be a powerful driver for new service creation if released as under open source licence.

**Box 3: Data and productivity**

To reframe the value of data in terms of increasing productivity, the questions might be stated as:

- Which data sets can be analysed to make an existing operation more efficient?
- What new services can be created by analysing existing data in new ways?
- Which new data sets could be created which would in turn create new services?

### 2.3 Accounting for value of data - the need for a framework

The economic argument that bigger, better data be considered a primary factor of production gains legitimacy when an accounting standard is developed to measure and test the value of data. The challenge is how to frame that accounting standard.

In an economic framework, data are treated in a similar way to intellectual property. The accounting treatment of intellectual property is however itself inadequate. There is currently no Australian Accounting Standard that comprehensively addresses the accounting treatment of intellectual property. IP Australia states that most relevant accounting standards include:

- AASB 138: Intangible Assets
- AASB 136: Impairment of Assets

IP Australia further states that many Australian companies do not even recognise their acquired intellectual property. Their intellectual property assets are instead often included on their financial statements as goodwill.

Framing the accounting treating of data as intellectual property becomes increasingly inadequate when the data are not generated by the entity which uses them. In the case of Twitter or Facebook,
user-generated data are harnessed to create services. These services encourage creation of more user-generated data and so more services. The productivity of the platform, and so the value of the platform to the business, is driven by ever more users freely contributing data.

Once captured, the question often arises as to who owns the data generated. Do users own the data they generate or are data owned by the entity providing the service? In some cases it can be said to be owned like any intangible asset. Twitter for example offers licences to different volumes of aggregated Twitter feeds including the full ‘Firehose’ (Twitter 2015). Thomson Reuters will also sell different volumes of current and historical market data on a commercial basis. In other cases, attempts have been made to control ownership of personal data. Famously Jennifer Lyn Morone tried to regain control of her data by becoming Jennifer Lyn Morone™ Inc (JLM), registered in the US state of Delaware (the Economist 2014).

The data licenced by Twitter and Thomson Reuters can be described in terms used to describe intellectual property: as an identifiable, intangible asset which has an indefinite useable lifetime. Further, it can be argued that the data do not suffer impairment with each new licence sold similar to intellectual property manifest as software. However when dealing with bigger, better data as described above, data are different to other intellectual property.

Data are unlike a brand in that they can be used in many different ways for many different applications. They are different to software in that they may not be produced (or purchased) by the entity that uses them. Data are different to goodwill in that they can be sold in discrete volumes to individual users. They are different to copyright in that they needs resources to capture, contain or transport them. Data are even different to community generated intangible assets such as open source software in the sense that a company, such as Facebook, can uniquely own the collection of data created by many individuals.

With the challenges associated with recognising the value of data in an accounting sense, it is not surprising that data are undervalued as a factor of production in the digital economy.

3. Some considerations when applying data and analysis

Before a discussion on data as a factor of production, it is worth briefly discussing a number of common conceptual challenges related to the use and analysis of bigger better data: simplification, ‘noise’ and the question of how many data are enough.

3.1 Simplify at your own risk

It is a truism that the more factors involved in a decision, or the more information available, the harder it is for human beings to make sense of it. Simplifications are often applied to help people make sense of data. An appreciation of more advanced statistical models and approaches can help immensely.

One of the most powerful simplifications used is the well-known ‘central limit theorem’. In mathematical terms, the theorem states that the distribution of the sum (or average) of a large number of independent, identically distributed variables will be approximately normal, regardless of the underlying distribution. The impact that the central limit theorem has had on popular thinking is
It is a commonly held belief that every random process can be reduced to a normal distribution, uniquely described by a mean value and a standard deviation. In practice, the requirements of variables being independent and identically distributed are overlooked, and that the outcome is only ‘approximately normal’ is ignored. When such a simplification is produced from billions of transactions or trades, ignoring the exact detail of a distribution can be disastrous. The concept of ‘black swan’ events was popularized by the writer Nassim Taleb in his book *The Black Swan: The Impact of the Highly Improbable* (Taleb, 2008). The essence of his work is that the world is severely affected by events that are rare and difficult to predict. The implications for markets and investment are compelling and need to be taken seriously.

**Box 4: Rarity and the central theorem simplification**

In the context of the central limit theorem simplification, we can artificially create a false view of rarity. For a normally distribution discrete random variable, a ‘rare’ event is one which is many standard deviations away from the mean. For example, a normally distributed discrete random variable has a probability of 0.000000000256 per cent of producing a value exceeding the mean by seven standard deviations, or once in every 390,682,215,445 events on average. For an event sampled at microsecond rates, such as market trading, a random variable will exceed this level approximately once every four days on average. When the distribution is only ‘approximately normal’, this thinking can be very misleading. A random variable with a slightly different distribution, with a thicker tail, can produce many more values far from the mean than the normal distribution simplification allows. Even with 10 times as many events at seven standard deviations from the mean, the normal distribution would still appear to be a good approximation. When we artificially brush away the probability of highly unlikely events, we do so at our peril.

**3.2 Unexpected correlations, human bias and noise**

The data produced by the world’s markets are hundreds of gigabytes per day, leading to a historical record measured in petabytes (Rozetta 2015). When faced with such enormous volumes of data, businesses can struggle to distinguish between meaningful causation and spurious correlations. One of the biggest issues is that of ‘noise’, where the incidence of random artefacts becomes so great that it obscures the underlying patterns that could be of use to researchers and businesses. Noise can also randomly generate apparent correlations where there are none, or potentially lead to false positives.

**Box 5: Distractify.com**

A US website highlights interesting correlations between variables which seem exceedingly improbable. These include the strong correlation between the number of honey producing bee colonies in the US and the marriage rate in Vermont (94 per cent correlated over the period 1999 - 2009), the total number of political action parties in the US and the number of people who died by falling out of their wheelchairs (approximately 92 per cent correlated over the same period), or

---

4 Random variable are independent and identically distributed if each variable has the same probability distribution as the others and all are mutually independent.
between the US per capita consumption of margarine and the divorce rates in Maine (>99 per cent correlated over the same period).

Source: Distractify 2015.

Whilst it may seem easy to dismiss a causal link between margarine consumption and divorce rates, many subtle interactions exist in markets, society and the broader economy. Anticipating a correlation between the closing price of the Australian dollar and the price of gold is far more plausible, as may be finding a correlation between property prices and auction clearance rates in a particular neighbourhood, or the success of a beer marketing campaign and the weather in a particular area.

In all of these examples, human bias about expected correlations can lead to a dismissal of potentially critical insight and even poor decision making. Despite the humorous nature of the example, for at least a decade there was a correlation between margarine and divorce rates in the US. The connection between gold and the Australian dollar may be weak, may vary significantly over time and may not be enduring. Figure 2 shows the Spot Gold Price versus AUD/USD currency pair over time and corresponding degree of correlation. The correlation varies widely over the period shown, ranging from being uncorrelated (close to 0 around August 2010), to be negatively correlated (close to -1 in early September 2010), to being highly correlated (close to 1 from October to November 2010).

Figure 2. Spot gold price versus AUD/USD over time and corresponding degree of correlation

Source: Roy 2010.

By relying more objectively on data to validate correlations and less on instinct or explicit models, human bias can be minimised in decision making. Applying independent data sets to test an apparent relationship using second, third or even higher order correlations can build greater confidence in a perceived causal link.

Using relatively simple mathematical techniques, the correlations $r_{x_n y}$ between predictor variables $x_n$ (independent variables such as margarine consumption) and a target variable $y$ (dependent variable such as marriage rates), can be used to produce the correlation matrix $R_{xx}$ of
inter-correlations between predictor variables. If the variables are uncorrelated, \( R_{xx} \) is an identity matrix\(^5\). In practice, some non-zero correlation will exist between any finite number of samples taken from a discrete random variable in real-world systems. These non-zero correlations do not reflect an actual correlation but are a consequence of practical system limitations creating ‘noise’ in the correlation matrix. As the number of predictor variables examined increases, the size of the correlation matrix \( R_{xx} \) increases, and the ‘noise floor’ of spurious non-zero correlation values increases. This limits the number of predictor variables which can be used for validating correlations.

As the noise is generated by finite sampling durations, finite precision numerical analysis and practical system limitations, sophisticated signal processing and numerical analysis techniques can be applied to estimate and characterise the noise floor. This in turn helps ‘amplify’ the actual correlations from the noise and so minimise spurious correlations. This is the subject of a great deal of active research in financial markets (Burda and Jurkiewicz 2008) and other areas (Burda et al. 2008). Balancing the trade-off of the number of variables explored depends on the sophistication of the modelling, the amount of computing resources available and the cost of computing the outcome.

3.3 The strength of distant data

In his highly influential 1973 paper, *The Strength of Weak Ties*, Mark Granovetter argues that in marketing, information science, or politics, weak ties enable the reaching of populations and audiences that are not accessible through strong ties (Granovetter 1973). The paper also highlighted the importance of linking micro-level interactions to macro-level patterns.

The bigger, better data equivalent might be rephrased as secondary data can be used to provide insight into a system or process for which primary data are not available. If the data you need to answer a question are not available (set A), it is possible to build a portfolio of information from related data sets (sets B,C,D) which gives insight to the answer you would have obtained from set A. The difference between ‘data’ and ‘insight’ is analogous to linking micro-level interactions to macro-level patterns.

### Box 6: The Department of Human Services

The Federal Department of Human Services (DHS) in Australia is one of the largest departments in the Australian Public Service. The 2013-14 annual report states that the department delivered $159.2 billion in payments to customers and providers, it operates 5 million online accounts, has a smartphone application which carried out 36 million Centrelink transactions, and almost 100 million online transactions between Centrelink and Medicare (DHS 2014). The DHS knows a great deal about who is doing well in Australian society and who is experiencing economic and social hardship. DHS would like to be able to identify are those people in between these extremes who are sliding down the socio-economic scale. These downward movements are often triggered by events such as the loss of a job, a serious injury, or the death of child or a partner. There are good reasons for wanting to know if someone is experiencing a trigger point. Interventions are typically more effective if applied earlier, whether it be emergency housing, a direct payment, management of bill payments,

---

\(^5\) The identity matrix or unit matrix of size \( n \) is the \( n \times n \) square matrix with ones on the main diagonal and zeros elsewhere.
counselling or financial planning assistance.

Gaining direct information about an individual experiencing a trigger point is difficult. Gaining insight that someone may be beginning to experience hardship from a their interactions with the DHS through call centres, online DHS account activity, email, use of social media and other digital sources can dramatically increase the ability to detect if someone is experiencing a trigger point. Through a multi-year alliance with CSIRO, the DHS carried out a portfolio of projects to build models based on available data to help answer the question of who is experiencing a trigger point and to better target delivery of services (CSIRO and DHS 2014). The projects drew on a wide range of data from natural language processing to geospatial modelling and behavioural economics. Its greatest unforeseen outcome highlighted by the DHS was ‘insight into the department’s processes and structures which were once thought too complex to understand, let alone alter’.

3.4 Variety – how wide to cast the net?

The principal explored in the DHS example in Section 3.3 is simply: the more related data sets of different types that you can bring to bear on a problem (sets B, C, D), the more completely you can build the portfolio of information which provides insight to answer the desired question. If you have the data you need to answer the question (set A), then drawing additional data sets allows you at ask more fundamental questions.

In financial markets, the term ‘implied volatility’ is the estimated volatility of a security's price. In addition to known factors such as market price, interest rate, expiration date, and strike price, the implied volatility is used in calculating an option's premium. The implied volatility for an option can be derived using a parametric model such as the Black-Scholes Model⁶.

Given sufficient computing resources, time and market data, it is theoretically possible to calculate the implied volatility of all securities in any given market at any point in time. A moment later, any one of the parameters associated with the volatility may change; the market price of the security may change in response to a trade, interest rates may change, or, in the case of an index, several constituent securities may change price. By knowing the correlation (or covariance) of any given security with all other securities traded, or by knowing the correlation of all securities with the movement of the market as a whole (the Beta of each security), an investor could make a very well informed decision about how to trade an option or the underlying security.

The data needed to gain the market insight and to make trading decisions are available and, importantly, they are available to everyone trading in the market who is willing to pay for it. Ignoring for a moment the accumulation of the correlation ‘noise floor’ described in Section 3.2, this becomes a brute force compute problem. Whoever has the greatest computing power should be able to produce the most complete and reliable solution for the next trade.

⁶It is worth noting that the Black-Scholes pricing model is based on a normal distribution of returns from the underlying securities, which implies that the underlying securities prices themselves are log-normally distributed.
The complexity of the problem however is so great that it is almost certainly infeasible to undertake an exhaustive computation, calculating all possible correlations for all securities and implied volatilities for all options. Insight may be gained by bringing other data sets to bear on the problem (sets B,C,D) even though the data needed to address the immediate issue is available (set A).

Whilst markets operate at micro-second rates, data sets associated with human rates of response may help frame the range of longer term outcomes, providing a broad trajectory for the movement of a security’s price. The extremely widespread use of social media means tools monitoring sources such as Twitter are often able to identify major events before mainstream media or government agencies. Twitter has been involved in the early detection of earthquakes\(^7\), fires\(^8\) and other natural disasters alerting other Twitter users of events before government agencies or mass media become aware. The early knowledge of a natural disaster can dramatically change a trading strategy.

**Box 7: Emergency Situation Awareness**

Researchers at Australia’s CSIRO have developed novel approaches to identify, track and understand emergent issues in massive social media data sets (CSIRO 2015). The Emergency Situation Awareness (ESA) software exploits the statistical incidence of words used on Twitter to describe emergency events. The tool is trained using historical word occurrences from past disaster incidents allowing ESA to reveal emerging topics and flag them for investigation. Accurate analysis requires combining natural language understanding, machine learning, data mining techniques, text classification, online clustering and geo-tagging. These techniques are adapted and optimised to deal with high-volume, real-time text streams to identify early indicators of unexpected events, explore the impact of identified incidents and monitor the evolution of events.

ESA has been used in front line environments for near real-time alerting of emergencies – including fires, floods and earthquakes – gauging community response to an emergency warning, and as a tool for gathering evidence of pre-incident activity. ESA has been used to help agencies understand the scale and fine grained impact of emergencies and disaster situations.

Social media may also be used to gain fine grained sentiment information following a major company announcement, such as change of CEO or a major cyber-security breach, which again may affect the trading strategy for major public companies.

**Box 8: Sony Corporation**

On December 17\(^{th}\) 2014, Sony Corporation in the US suffered a major cyber-attack widely believed to have been instigated by the North Korean government in response to the planned release of a controversial new movie, *The Interview* (Yadron et al. 2014). The complexity of the situation was unprecedented – Sony is a widely traded, publically listed company, a leading provider of electronic consumer goods, a major figure in the entertainment industry. Further, the attack was speculated to be politically motivated, and the attacker was a nation state known to be hostile to the US. These myriad factors made understanding the public response and anticipating the market response

---

\(^7\) See for example early detection of 2014 earthquake (Prigg 2014)

\(^8\) See for example FAAST fire detection from Honeywell [https://twitter.com/faast_detection](https://twitter.com/faast_detection)
extremely difficult. A US government response was broadly anticipated and reported in mass media. A fine grained understanding of consumer and broader public sentiment gained from social media analysis could have helped inform trading strategies for securities based on Sony in the immediate aftermath of the reporting of the attack. Following the announcement of the attack on December 18th, the share price fell slightly on 19th December, then recovered in the coming days as shown in Figure 3.

Figure 3. Sony share price on the NYSE around the time of the December 2014 Cyber-attack

Source: Sony.

4. Coming to grips with the digital economy – moving bits not atoms

The term ‘Digital Economy’ was first used in Don Tapscott’s 1995 book, *The Digital Economy: Promise and Peril in the Age of Networked Intelligence* (Tapscott 1997). Nicholas Negroponte subsequently used the metaphor of shifting from ‘processing atoms to processing bits’ as a means of explaining the transition to a Digital Economy. The Digital Economy is now described by some economists as the sector of the economy associated with zero marginal cost intangible goods available over the internet (Fournier 2014).

Digitisation of commerce, goods and services has had a dramatic impact on all industry sectors and has led to the demise of household names who failed to adapt. Famous examples include Blockbuster video, Borders books and Eastman Kodak. The power of digitisation in commerce is threefold:

- the dramatic reduction in the marginal transaction and delivery costs of digital goods and services;
- the reduced costs to consumers of access, discovery, and comparison of goods and services;
- the ability of service providers to develop extremely high levels of customer intimacy based on ‘metadata’ generated around search, purchase, shipment, use, user experience and feedback of digital products and services.
Supplier generated, user generated and system generated data play a major role in these areas.

4.1 Digitise

Anything which can be digitised can be delivered as a digital file. The interest in 3D printing is driven by the ability to model and map physical objects in digital form and then send anywhere for recreation. The recent example of a NASA astronaut using a 3D printer to create a spanner on-board the orbiting space station, based on a design sent on demand from Earth, hints at how far this technology could extend (BBC 2014).

The increasing dominance of digital services is increasingly making the world a single global market place. A recent McKinsey Global Institute report shows that, in 2012, flows of goods, services, and finance reached $26 trillion, or 36 per cent of global GDP representing a 50 per cent increase (relative to GDP) compared to 1990 (McKinsey Global Institute 2014a). If the spread of digital technologies and rising prosperity in emerging economies continues, global flows could nearly triple by 2025. As global flows of services has increased, their direction has also changed. Developed economies once dominated global trade with 54 per cent of all goods traded in 1990 between developed economies. In 2012 these flows accounted for only 28 per cent.

Underlying these flows is the dramatic increase in exchange of data and communication across borders. Global online traffic has grown from 84 petabytes a month in 2000 to more than 40,000 petabytes a month in 2012, a 500-fold increase. Cross-border voice traffic has more than doubled over the past decade, primarily due to digital calls. In 2012, Skype call minutes grew to 39 per cent of the level of traditional international phone call minutes, an increase of more than 500 per cent since 2008 (Economist Intelligence Unit 2015).

Figure 4: Global flow of goods, services and financial flows; share of GDP


Figure 5. Share of selected cross-border flows that are digital per cent in 2005 and 2013.
Once delivered, digital goods and services can of course be used in a ‘digitally equivalent’ way to their non-digital counterparts. In addition, data associated with the end users’ interaction, feedback and use experience can also be captured by the producer to improve the offering and allow for greater personalisation of experience for each user. A new application downloaded to a phone or laptop will request access to location information, ask to allow use statistics to be sent anonymously to the provider, and even ask a few questions about end user preferences. These requests are often associated with a benefit offered to end users of additional functionality from the applications; restaurant suggestions based on learned preferences, driving or walking directions based on current location, news based on personal interests.

4.2 Link

Creating value and insight starts with digitisation of services and gathering data in searchable, electronic form. Linking it together is the next step.

The vast amounts of data that can be collected from systems in a business, and the interaction of users with those systems, has long been the raw material which underpins business process reengineering. Collected data can be used to highlight challenges, identify inefficiencies and feed models for return on investment. A step change can be realised by linking and analysing all of the data from all the systems in a business, and all of the user interactions with all of those systems.

Figure 6 shows a simple generic data architecture focussed on collating and transforming data (data layer) to provide input to operational management, performance management or configuration management (information layer), which in turn feed domain specific applications (knowledge layer).

**Figure 6. Simplified Data architecture**
4.3 Model (and analyse and optimise)

Much of the growth of customer experience management or customer centric business has been driven by inverting the traditional view of business. Moving the viewpoint from ‘multiple users interacting with a monolithic business’, to ‘modelling, analysing and then optimising the interactions of individual users (customers) with the major points of interaction with a business’.

Customer-centric businesses collect, link, model and analyse data on customer use history, engagement with call centres, billing systems, help desks, complaints processes and many other points of interaction, developing one view of the customer experience. The deeper insight generated from these data allows services to be targeted at an individual level, allows identification of high revenue and high cost users, and helps inform where scarce resources should be applied to maximise outcomes.

By extending this approach to collecting, linking, modelling, analysing and then optimising all the touch points of users, employees, customers, patients, contractors or subscribers with all of the systems, the customer centric model can extended to develop multiple ‘centres’. The organisation can become employee centric, patient centric, doctor centric, provider centric and even new -business centric depending on need. The reproducibility of data for multiple purposes means that all views could operate simultaneously.

Box 9: Comptel

Finnish company Comptel develops software to create a customer centric view of mobile networks. An average mobile phone generates 10s of megabytes of data per day which is not user calls or user data. These data are used for connection to the network and ensuring network quality. These data have historically been used to optimise overall network performance and coverage quality, taking an operator centric view of the performance information.

The Comptel software uses a wide range of non-user data (call fail rates, call drop rates, call quality, call duration patterns) as well as user billing information, to predict the likelihood of an individual user ‘churning’ or moving to another mobile network operator. Comptel claim to be able to identify 75 per cent of potential churners. Identification of potential churners allows operators to analyse
the ‘value’ of the customer and intervene to keep the customer if warranted. This insight creates substantial financial benefits for the operator.

Source: Comptel 2015.

4.3 Personalise

Gaining insight about preferences is a constant focus for digital services companies. Inferred from retail purchases or credit card transactions, based on web page visits, or gathered explicitly from inputs into recommendation systems, every time a customer accepts or declines an offer, more preference data are collected.

Great value is created for the application provider when individual users provide explicit information about location and interests. One of the most famous examples occurred in 2012 when the supermarket chain Target predicted a teenage girl was pregnant based on the spending patterns recorded through her loyalty card, information explicitly shared with the store (Forbes 2014a). The downfall for Target was that they sent purchase recommendations for a pregnant woman direct to the teenage girl. This led to complaints from the girl’s father protesting Target’s actions. Target was correct. In Australia, Woolworths uses purchasing information gained from loyalty card data to provide individualised ‘specials’ to their customers. This requires an explicit opt-in agreement from each customer.

Applications providing feedback, even anonymised, allow creation of models of implied preference which creates value for the application provider. Analysing data such as which restaurant suggestions are explored further, which news sources are accessed when and how often, as well as user location can be used to create models of individual preferences, socio-economic status and even gender. In their 2014 paper, researchers analysed approximately 7 million users and over 1 billion communication records (call data records and SMS) to develop fine grained demographic information and to uncover interesting social patterns (Dong et al. 2014). This in turn can lead to much more finely targeted recommendations for individuals, or assist with broader planning and service provision.

Box 10: Extreme personalisation

Since 2001 the Human Genome Project sequenced the first human genome at a cost of USD $100 million, the cost of genome sequencing has fallen dramatically to approximately USD $1,000 by late 2014. Full human genome sequencing can provide raw data on all six billion nucleotides in an individual’s DNA in the form of a string of four symbol codes, effectively representing the basic building blocks of a human being as data.

US Company EdicoGenome has reduced the ‘wet chemistry’ of genome sequencing to processing on a single chip, producing up to 250 gigabytes of data per individual, followed by dedicated intensive data processing on dedicated hardware processor. EdicoGenome claims to be able to sequence an individual’s genome in minutes.

As yet, the raw genome data do not directly allow individual design of drugs to treat individual disease. The data can however be fed into analytical models which can be used to predict
susceptibility to certain hereditary diseases and to determine the most appropriate class of drugs to treat an illness for the individual. The future of highly personalised medicine relies on the ability to process much more of the genome data and experiment with drugs which more accurately match the exact requirements of the individual, their illness and the impact of the delivery mechanism in the body.

Source: EdicoGenome 2015.

4.4 The changing nature of work

Digital technology is also changing the nature of ‘working’. The ubiquity of email has changed the nature of remote work or working across multiple sites. Low cost communications tools such as Skype video conferencing, or more formal collaboration environments such as Atlassian’s JIRA or Confluence, continue this trend (Atlassian 2015). There is also growing online labour market, such as Freelancer, which connect tasks and people from around the world (Freelancer 2015). Movement of people is being replaced by movement of digital files, often over high speed connections, to provide the illusion of human real-time interaction.

Taking a step beyond a digitally replicating an office work environment, bringing additional data to bear means the experience can be enriched significantly. Workers in a call centre can be supplied with trending social media issues related to a product or service, individual user experience information or even user preferences.

An employee also creates a substantial digital footprint within an organisation. One of the current hot topics in human resources is the ability to analyse data associated with that footprint to determine their likelihood of leaving an organisation.

Box 11: Staff retention

US company Ultimate Software Group assigns clients’ employees, and even its own workers, individual ‘retention predictor’ numbers, similar to a credit score, to indicate the likelihood that a worker will leave in the near future (Ultimate Software 2015).

VoloMetrix, also from the US, examines HR data as well as anonymized employee email and calendar data, claims to predict flight risk up to a year in advance for employees who were spending less time interacting with colleagues or attending events beyond required meetings (Volometrix 2015). With an estimated median cost (US) of turnover for most jobs of around 21 per cent of an employee’s annual salary9, and an estimated average cost $3,341 to hire a new employee10, this analysis can provide an immediate return on investment for companies.

5. The changing financial sector

In many parts of the developed world there has been an acknowledgement that the recent financial crisis and its aftermath have challenged beliefs about the way financial systems, businesses and households behave. Customers are demanding greater personalisation of services, greater flexibility

---

9 Estimated from the Center for American Progress, 2015.
in access, and greater transparency. At the same time, they are demonstrating a higher propensity to change financial service providers or use non-traditional sources of finance, savings or retirement products. Regulators have sought ways to better protect customers, improving the resilience of the financial system, and better ways to identify potential for major disruption.

In 2015, the Bank of England released a discussion paper which called for the use of new data, methodologies and approaches to understand household and corporate behaviour, the domestic and international macro-economy, and risks to the financial system (Bank of England 2015). The Bank posed questions as to how to best take advantage of an increasingly wide range of data sources and analytical tools to improve understanding of economic and financial behaviour. It also reflects greater awareness of the advantages that come from using micro-data, better access to administrative data held by authorities, and technological improvements that have lowered the cost of online surveys.

In Australia in 2013, the Federal government launched an independent Financial System Inquiry (FSI) to examine how the financial system could be positioned to best meet Australia’s evolving needs and support Australia’s economic growth (Treasury 2014). The FSI was tasked with examining issues such as technological change, Australia’s global competitiveness and offshore regulatory frameworks. In late 2014, the FSI released a report with recommendations to promote a competitive and stable financial system that contributes to Australia’s productivity growth and promoting the efficient allocation of capital and cost efficient access and services for users.

Australian banks and financial sector firms are heavy investors in new technology. Following the lead of the major banks which have invested in improved ICT and data sharing systems, many organisations are improving their organisational productivity and creating new and valuable services for their customers. In terms of productivity improvements, broad areas for investment include:

- Effectiveness and efficiency – connecting the different internal views of the bank’s business by joining data sets, for example enterprise with retail, to improve the way banks operate.
- Old things in new ways – gaining a more individual understanding of customers’ preferences and propensities by doing more with transactional information, including credit card transaction information.
- New things in new ways – delivering new services by linking data from partners in a trusted data ecosystem in near real time.

Banks continue to invest heavily in system resilience, cyber security and regulatory compliance, however these are not considered in detail in this paper.

**Box 12: The Commonwealth Bank**

The Commonwealth Bank in particular has received praise for building on its investment in the SAP core system, for being the first bank to perform property purchase settlement on the newly established Property Exchange Australia platform and the implementation of e-documents for new home loan customers (PEXA 2015). The bank has also developed an application interface for their customers and claims more than 3 million unique users who collectively are logging on 15 million
times per week, carrying out $2.5 billion of transactions (Pearce 2015).

5.1 Levelling information asymmetry – innovate or get disrupted

McKinsey & Company describes the digitisation of banking and finance as an ‘irresistible trend’ powered by the reinforcing drivers of connectivity, automation, decisioning (driven by big data and enterprise data) and innovation (McKinsey & Co 2015).

This trend is changing the historical nature of the relationship between financial services companies and their customers. Due to their scale, institutional bodies such as banks, trading houses and insurance firms have traditionally had greater access to far deeper data than individuals, creating ‘Information asymmetry’ or ‘knowledge gaps’ between institutional bodies and individuals (retail investors or the average person). This information asymmetry creates the opportunity for institutional bodies to capitalise on inefficiencies in factors like pricing.

As more customers move to online digital banking, the power of internet search, the enhanced ability to compare, and a rising expectation for personalised services are changing customer behaviour in finance as with every other sector. KPMG’s 2014 report on the Sydney FinTech Sector highlighted that modern finance customers are increasingly likely to be ‘more demanding, less loyal, more likely to take advice form peers, more willing to adopt new technology, operate in a world of reduced information asymmetry and will insist on highly personalised services’ (KPMG 2014).

Against this background, a growing number of start-ups and smaller enterprises are capitalising on these trends and challenging the traditional information asymmetries by using data and analytics to expose inefficiencies. Some early entrants including software giants and telecommunications companies sought to replicate the whole banking process, resulting in business models that really only appealed to the most tech-savvy or price-conscious customers. Many of today’s innovators are aggressively targeting areas of high frustration for customers and high profitability for incumbents, eroding incumbents’ most valuable products.

5.2 Know your customer, know your product (then join the dots)

Customer data are a central decision-making factor for financial institutions. Banks make lending decisions based on individual credit scores and insurers may look at a driving record or require a health check before issuing a policy.

Many banks have developed accurate models of credit risk based on purchasing, borrowing and payment history. The use of online or application based interfaces between banks and customers helps to increase the amount of data and the variety of information banks understand about their customers. If the bank can gain insight into personal preferences, and can link with data partners in its ecosystem, for example residential property information, a powerful potential is established. Banks could recommend properties to their customers which meet their general preferences and with pre-approved loan terms, before the customer has even seen the property. A net positon can be presented to the customer along with ‘accept / decline’ option. Add this property to the portfolio or wait for the next option. Whilst safeguard need to be built into a system such as this, the potential exists to dramatically reduce the transaction costs and the time involved in buying property for investment purposes.
For this new business model to work, a confluence of risk analysis, property value and investor preference information is required. Components of this confluence exist already. A new breed of insurance company is identifying ways to generate streams of data that help them make better pricing decisions and encourage their policy-holders to make smart decisions.

**Box 12: Credit risk**

In Australia, Veda has a core product offering which includes the provision of credit reports for individuals and businesses. Veda’s data includes credit information on more than 20 million individuals and 5.7 million commercial entities in Australia and New Zealand (Veda 2015).

Australian company MARQ Services provides an independent risk assessment for the Australian mortgage loan funding markets based on a set of data requirements with clear definitions, specifying a comprehensive set of risk and other loan-level information for mortgages. Its data requirements cover characteristics of the borrower, characteristics of the security property, terms and conditions and performance of the loan and other information, such as the originator of the loan (MARQ 2015).

### 5.3 Peer-to-peer lending – a new business model

One of the major concerns of the established banks is peer-to-peer lending which targets the high volume, small denomination loans – a profitable market segment for institutional lenders. Peer-to-peer lending platforms are online financial communities where investors and borrowers can connect for small, unsecured loans. Using an online credit scoring service, these companies tailor rates to borrowers based on a highly personalised set of details.

One of the main advantages of peer-to-peer lending for borrowers has been better rates than traditional bank rates can offer. The advantages for lenders are higher returns than can be obtained from a savings account or other investments. The interest rates are set by lenders competing for the lowest rate on the reverse auction model, or are fixed by the intermediary company on the basis of an analysis of the borrower’s credit.

The first company to offer peer-to-peer loans anywhere in the world was Zopa (Zopa 2015). Since its founding in 2005, Zopa has issued an estimated £500m of loans and is currently the largest UK peer-to-peer lender with over 500,000 customers. In Australia, SocietyOne, DirectMoney and RateSetter are examples of peer-to-peer lending platforms. Somewhat like social networking sites, these are online financial communities where investors and borrowers can connect for loans of up to $35,000 (SocietyOne 2015; DirectMoney 2015; Rate Setter 2015).

Borrowers assessed as having a higher risk of default are assigned higher rates. Lenders can mitigate the individual risk that borrowers will not repay loans by choosing which borrowers to lend to, and mitigate total risk by diversifying investments among different borrowers. The lender’s investment in the loan is however not protected by any government guarantee.

As people and their devices become more interconnected, new streams of granular, real-time data are emerging, and with them innovators who use that data to support financial decision-making. In
the peer-to-peer case, borrowers can use online services to build and even advertise their credit score.

**Box 13: FriendlyScore**

UK company FriendlyScore analyses people’s social networking patterns to provide an additional layer of data for lenders trying to analyse the credit-worthiness for a borrower. At the time of writing this article, the author’s ‘FriendlyScore’ of 61 per cent, based on analysis of Facebook and LinkedIn profile rated the author as good risk for a loan. Adding Twitter brought the score to 70 per cent. A ‘perfect’ candidate for a loan based on categories of education, work, credibility and network. FriendlyScore also makes it possible to share the outcome on your social media networks, putting the individual at the centre of the relationship with lenders.

**Figure 7. FriendlyScore based on Social media analysis.**

![FriendlyScore](image)

*Source: FriendlyScore 2015.*

**5.4 Crowd funding – extending the peer-to-peer concept**

Crowd funding takes the concept of peer-to-peer lending to a new level representing a further threat to traditional debt funding for companies. Crowd funding gives access to potentially significant levels of financial capital to fund a project or venture by consolidating small contributions from a large number of people.

The two main models are the ‘keep-it-all’ model where the entrepreneurial firm sets a fundraising goal and keeps the entire amount raised regardless of whether or not they meet their goal, and ‘all-or-nothing’ model where the entrepreneurial firm sets a fundraising goal and keeps nothing unless the goal is achieved (Cumming et al. 2015). A more controversial version of crowd funding is equity crowdfunding where the backer receives shares of a company, usually at an early stage, in exchange for the money pledged.
Crowd funding has the advantage of allowing entrepreneurs to pre-sell a product or service to launch a business concept without incurring significant debt or sacrificing equity. It also has the advantage of providing direct market feedback as to the potential success of a new project before that project has commenced.

Online crowdfunding sites have become leading providers of funds for start-ups, new concept development and individual projects ranging from thousands to millions of dollars. US Company Kickstarter famously helped the makers of the Pebble Smartwatch to raise more than $20 million in 2015 (KickStarter 2015; Chang 2015).

In late 2014, HSBC estimated that the 2013 market for crowdfunding was in excess of $5 billion globally (Forbes 2014b). In 2015, analyst firm Massolution predicted that the market would grow to more than $34 billion (Crowdsourcing 2015). UK-firm, The Crowdfunding Centre, estimated that between July and December 2014, more than 83,000 projects were initiated through global crowdfunding.

Crowdfunded finance is, however, still regarded as niche by many businesses. In March 2014, the UK Government consulted on whether and how it could legislate to create a mandatory process, whereby Small and Medium Sized Enterprises that have been rejected for finance are ‘linked up with other lending opportunities from challenger banks and alternative finance providers’ (UK Parliament 2015). As part of the consultation process, the British Chambers of Commerce (BCC) reported that 49 per cent of their members relied on banks or building societies for debt funding, 10 per cent used equity and 8 per cent used ‘grants, venture capital, private equity, peer-to-peer lending and angel finance combined’. The BCC concluded at the time that there was little understanding of alternative finance options amongst their membership.

In Australia, the 2012 guidance from the Australian Securities and Investments Commission (ASIC) stated that ‘crowdfunding as a discrete activity is not prohibited in Australia when raising funds with donations’, however provisions of the Corporations Act need to be considered if raising funds with either debt or equity (ASIC 2012).

In May 2014, the Corporations and Markets Advisory Committee released a report on equity crowdfunding proposing a regulatory regime to facilitate crowd sourced equity funding in Australia (CAMAC 2014). The report recommended the federal government introduce legislation allowing retail investors to invest up to $10,000 a year in start-ups through equity crowdfunding, with a maximum of $2,500 in each company. It suggested companies be allowed to raise up to $2 million per year on such platforms.

As part of the 2015 Budget, the Federal government announced that it would make it easier for small businesses to access capital by allowing crowd-sourced equity funding and by simplifying related reporting and disclosure requirements. As at time of writing, details on the specific model have not been released, however Treasury has set aside funds to enable ASIC to implement and monitor the regulatory framework to facilitate the use of crowd-sourced equity funding when it is released.
Box 14: VentureCrowd

Australian start-up VentureCrowd is a crowdfunding platform including equity crowdfunding, property crowdfunding and debt-based crowdfunding (VentureCrowd 2015). In May 2015, the company raised $4.2m for taxi-booking and payment software company Ingogo (Ingogo 2015). Whilst small compared to raising in the US, this represented the largest Australian equity crowdfunding raising to date.

5.5 Robo-advice – automating and personalising analysis of financial data

High quality, highly personalised financial advice is typically associated with institutional investors and high net worth individuals. Gathering the necessary information and distilling recommendations from that information is time consuming and requires skill, experience and judgement. Many firms offering financial advice will also rely on proprietary analysis models or unique data assets to differentiate their offerings and to charge premium prices.

A new group of innovators are using their technical skills to automate resource intensive processes such as financial analysis and advice. This allows them to offer services to a much broader cross section of customers including younger and less wealthy individuals. While customers must forego the in-person attention of a dedicated adviser, they receive many of the services they would offer at a lower cost. Robo-advice providers typically offer lower fees (ranging from 0.25 per cent - 1 per cent a year) and will take on clients with small portfolios.

Given the relatively recent advent of robo-advice, the question of long term effectiveness of human versus robo-advice is unanswered. Lance Roberts argues that there is an emotional mismatch between robo-advisors which take a long term view of investing, based on historical returns of markets over time (Roberts 2014). For many human investors, he argues that the reality is that their ‘long-term’ time frame is only from today until the next major market correction begins.

Nonetheless, robo-advice is becoming increasingly popular. For an individual investor with a small amount to invest, taking robo-advice allows them to access a wider set of data and market fundamentals than they are likely to be able to address themselves. Launched in December 2011, US company Wealthfront reached $2 billion funds under management in 3 years. It now has more than $2.5 billion under management (Wealthfront 2015). The Investment Trends August 2014 Advice Report showed that 42 per cent of the over 6 thousand Australian adults surveyed are self-directed investors who prefer to invest themselves (Santhebennur 2015). Additionally a quarter of respondents reported that they would prefer piece-by-piece advice if it is cheaper and that 10 per cent of Australian adults prefer to get piece-by-piece advice online, predominantly among those aged between 30 and 44.

Box 15: Robo-advice

In early 2015, Midwinter undertook a survey on robo-advice reaching almost 300 financial advisors in a wide cross-section of the industry. More than half of the advisers surveyed were aware of robo-advice and not concerned about its potential to disrupt the advice industry, with 12 per cent of these advisers actually excited about its arrival. Approximately half of advisers polled thought that
ASIC should intervene and limit the scope of robo-advice in some form including limiting robo-advice to non-product advice only, to super fund (formerly RG200) scope only, or to intra-fund advice only.

Robo-advisers in Australia such as InvestSmart have automated a wide range of wealth management services including asset allocation, investment advice and even complicated tax minimization strategies, all offered to customers through an online portal (InvestSmart 2015). InvestSmart uses digital technology to analyse revenues, assets and debts, then uses that data to make decisions optimised for the goals of the investor.

5.6 Feeding the robo-traders – big data for high frequency trading

One of the most controversial trends in modern financial services is that of high frequency, or algorithmic trading where machines make buy/sell decisions millions of times per second. The best-selling book *Flash Boys* provided an insight into the world of high frequency trading, some of the strategies of traders and possible consequences of these strategies (Lewis 2014).

Algorithmic trading is the continuous scenario planning and optimizing process (as described in Section 4.3), although in this case operating on a model of the market or the future value of a tradeable instrument. The simplest models are a near exhaustive set of ‘if-then-else’ conditions which consider discrete cases with a finite set of possible outcomes for a small number of tradable instruments. The most sophisticated algorithms consider a wide portfolio of instruments, factor in derived information on the portfolio of instruments (such as derived volatility described in Section 3.4), seek to identify and learn trends in the market through evolutionary or machine learning techniques, and trigger buy/sell decisions based on this learned behaviour.

High-frequency traders are typically considered to compete against other high-frequency traders, rather than long-term investors with firms making up the low margins per trade with very high volumes of trades. It has been argued that a core incentive in much of the technological development behind high-frequency trading is essentially front running, in which the varying delays in the propagation of orders is taken advantage of by those who have earlier access to information.

All trading algorithms are reliant on real time, high quality market data for their operation. As described in Section 3.4, the competition essentially becomes the ability to direct computing resources to the largest possible data set. The quality of the algorithm itself will be a trade-off between the sophistication of the modelling and analysis, and time taken to produce a result.

Historical market data is essential for the back testing of trading algorithms – evaluating trading decisions and comparing predicted versus historical outcomes. Within the course of a trading day, many millions of trading decisions may be made, so a wide variety of scenarios are explored by testing algorithms on days or even years of historical market data.

One major advantage arising from the ability to capture historical market data across many markets is transparency. Market data, available to researchers and regulators worldwide, provides a very transparent record of interactions in a market and across markets. This greatly improves ability to provide regulatory oversight and arguably improves the confidence of entities operating in highly liquid markets.
Box 16: Financial markets database

Sydney based not-for-profit SIRCA holds one of the world’s largest historical financial markets databases. Supplying the academic community and selected commercial partners, SIRCA provides millisecond-timestamped tick data dating back to early 1996. The database covers 45 million active over-the-counter and exchange-traded instruments worldwide, updates at a rate of more than 1 million messages per second, and is approximately 3 petabytes in size. The data are used by regulators, researchers and for back testing of trading algorithms.

SIRCA also maintains a comprehensive database of the Australian stock exchange tick history to 1991. The database includes equities, exchange traded options, warrants, every bid and ask order, every trade, including the volume of each order and executed trade, time and sales, market depth, intraday, end of day, and order book pricing.

Source: SIRCA 2015.

5.7 Investing for productivity growth – an unmet need for infrastructure data

It has been argued that Australia’s appetite for investment in housing has a negative impact on productivity growth in two ways (van Onselen 2014). Firstly demand for investment in existing property increases the price of land, one of the primary inputs for traditional businesses. Secondly, land held by rent-seeking investors, without transforming the land in any way, has a lower impact on productivity as compared to land transformed and used for production. The argument is that investment of excess capital that had has been directed into pre-existing housing would have greater impact on productivity if directed into businesses and infrastructure, as occurs in countries such as Germany.

The Association of Superannuation Funds of Australia reported that superannuation assets totalled $2.02 trillion at the end of the June 2015 quarter (ASFA 2015). Over the 12 months to June 2015, there was a 9.9 per cent increase in total superannuation assets. Despite this strength, the recent report from the FSI stated that:

A strong and efficient superannuation system is critical to help Australia meet the economic challenges of an ageing population. ... Superannuation assets are not being efficiently converted into retirement incomes due to a lack of risk pooling and over-reliance on individual account-based pensions.

This statement highlights a number of issues, one being Australia’s aging demographic. The number of Australians in the retirement age group is projected to more than double by 2054-55 compared with today, significantly increasing the number of retirees for every full time employee. This creates an imperative to increase productivity. A second is an over-reliance on a small number of investment options. Figure 8 provides a current overview of assets allocations for Australian pension funds and other investment funds. The overwhelming majority is in shares or similar equity.

Given the size of the funds under management, national productivity could be significantly impacted if superannuation funds could more widely invest in productive assets (capital) such as major infrastructure projects or commercial property. For the investor, infrastructure projects are long term investments that potentially match the long duration of pensions’ liabilities; infrastructure
assets linked to inflation could hedge pension funds liability to increasing inflation. Illiquid assets should also trade on noticeably higher yields relative to their more liquid counterparts. Finally, infrastructure investments have a low correlation to traditional asset classes and can provide diversification in an investment portfolio.

Figure 8. Assets of pension funds, life insurance corporations and non-money market investment funds

Source: ABS 2015.

The challenge for pension funds is the lack of objective, high-quality data on infrastructure investments and a clear and agreed benchmark. This makes it difficult to assess the risks of these investments and to understand correlations with the investment returns of other assets. Without such information, pension funds are reluctant to make such allocations. Undertaking the deeper analysis required to understand the risks and returns compared to liquid or listed assets typically leads to much higher management and holding fees.

What is needed is information on the expected performance of these assets including cash flow forecasts, debt repayments, performance and usage forecasts. Whilst this information relies heavily on assumptions and financial modelling, real world data can be used to help build confidence in these assumptions, in particular, the issue of usage forecasts.

Section 4.3 highlighted the study of more than 1 billion mobile call records as a way of gathering fine grained demographic information. Taking a step further to anonymously aggregate individual preferences and linking this to movement of people – journey start and end locations, journey times and dwell times – a model of ‘flow of preference’ can be developed for cohorts of people. The flow of millions of mobile devices mapped to fine grained (anonymised) demographic information becomes a powerful input to decision making. An understanding of flows of different demographic groups over time and space can strongly inform planning for a major piece of infrastructure, or the location of a new store.
Box 17: Streetlight Data

Gathering anonymised information from mobile device movement, US company Streetlight Data uses the principles of flows of preference to provide services to infrastructure planners, chain stores and governments about who goes where to enhance transportation engineering, urban planning, retail siting, and marketing decisions.

*Source: Streetlight Data, 2015.*

5.8 Assuring integrity by detecting anomalies

One of the best ways to detect fraud is to look for anomalies. Anything which stands out from ‘normal’ deserves closer inspection. This requires an understanding of what normal means. In any given set of data, those with the highest values or lowest values are obvious candidates, however with just one data set, it is impossible to determining whether these values are ‘reasonable’ or as expected. By linking related data sets together, it becomes possible to ask the question ‘is this value reasonable given the other conditions’. This is the basis of Bayesian analysis, a standard mathematical technique for a conditional hypothesis (Wolfram Research 2015). The more data dimensions which can be brought to the hypothesis, the more likely an anomaly will be identified allowing us to ask questions such as ‘is it reasonable that an expense was recorded on your credit card, when the previous purchases were several seconds before, in a different country, and you have no prior history of purchasing this type of good or service?’.

Doing this in real time for credit card users requires access to data sets associated with millions of users, their purchasing history and their assumed location, in near real time. It also requires creation of a model of normal behaviour before identifying a fraudulent transaction.

Box 18: Capital markets cooperative research centre

In 2011 there were more than 119 million non-referred (general practitioner) attendances processed by Medicare. The Capital Markets CRC has developed a health insurance fraud detection system by analysing data sets associated with hospital claims against the national healthcare system. Individual practitioners within a hospital use a unique procedure identifier code (IDC) to claim payment for carrying out that type of procedure (ACCD 2015; DoH 2013). The system is self-reporting in that practitioners select the procedure identifier and payments are (largely) made without independent verification. This can lead to ‘identifier creep’ where the identifier associated with a more expensive procedure is reported. This may be a matter of interpretation of the procedure by the practitioner (or their administrative staff), an accidental mislabelling of the procedure, or deliberate fraud.

The CRC is refining a rules based system to identify cases of identifier creep or deliberate fraud. The data sets the system relies on data from Insurance Funds including detailed claims data, hospital discharge data, member and membership data (selected information), provider data (selected information) and fund-specific benefits data. The system also uses the Commonwealth Medical Benefits Schedule code book, IDC codes, prosthetics data, diagnosis-related Groups and formulae.

*Source: CMCRC 2015.*

5.9 Personal data as a raw material - profiting from sharing
As more companies develop platform offerings which generate, and profit from, services based on aggregating user data, innovative companies are coming up with ways of sharing in the ‘monetization’ of user data, providing a platform to connect buyers for personal data to individuals in return for a brokerage fee.

Box 19: Data Republic

Dutch company Data Republic is a personal data brokerage company. The range of services includes secure storage of data, management of data sharing – allowing individuals to keep track of what data has been shared with whom, and will negotiate with prices on behalf of the individual sharing the data.

Source: Data republic 2015.

6. Digital transformation in government

Much like industry, all levels of government have the potential to benefit from increasing use of on-line digital technology and the associated use of data. The 2015 report, Digital Government Transformation, highlighted many benefits of digitising customer engagement with government. As one example, the report found that, of the estimated 811 million transactions at the federal and state levels each year, approximately 40 per cent are still completed using traditional channels (Deloitte 2015). If this figure could be reduced to 20 per cent over a ten-year period, productivity, efficiency and other benefits to government worth around $17.9 billion for federal and state governments.

Sharing data is highly relevant to the Federal Government’s National Digital Economy Strategy (Liberal Party of Australia 2015). One goal highlighted was ‘Four out of five Australians will choose to engage with the Australian Government through the internet or other type of online service’. Australia is currently nowhere near that level. Whilst progress has been clearly made, one recent estimate put the level at one in eight Australians aged over 14 (Roy Morgan Research 2015; Crucial 2015).

6.1 Transforming government through data

Major challenges identified to achieving the federal government’s National Digital Economy Strategy include concerns about government sharing of data:

- unambiguous authentication of citizens accessing services
- the quality of online services provided by government
- the ability of government to share data between agencies
- the willingness of the general public to the accept the concept of government agencies sharing data
To progress the digital transformation agenda, in January 2015, the Federal government announced the formation of the Digital Transformation Office (DTO). The DTO’s mission is to create digital services that are easy to find and use for both citizens and for different government agencies. Linked to transformative activities in other major government departments including the Australian Taxation Office and DHS, the DTO’s goals align to focus on improving the productivity of government, improving existing service quality for citizens and offering a new range of services to citizens enabled by digital technology.

Sharing of massive quantities of data between agencies is fundamental to the goal of creating the ‘citizen centric’ perspective with a single point of entry to government, and a single view of the experience of dealing with government. The DTO will attempt to develop a ‘whole of government’ view with the intention of improving the experience of dealing with government. Similar to that of industry, analysing every point of contact with every person who uses these government services provides the opportunity to create dramatically improved service quality, and also allows the personalisation of services expected from private sector service providers.

**Box 20: Digital Transformation Office**

The DTO will use technology to make services simpler, clearer and faster for Australian families and businesses. ... One of the DTO’s first tasks will be to ensure people no longer have to complete separate log on processes for each government service. Instead, people should have a ‘digital identity’, which they can use to log in to each of their services across the government.

[The DTO]... want data to be accessible to agencies in a machine-friendly API format, to integrate into their own reporting and business intelligence purposes so they can also make improvements. This may supplement agencies’ own reporting and, ... could reduce the scope of data agencies need to collect themselves. .... We are going to separate our data collection from the analysis and reporting tools and take an API-driven and modular design approach. This way we can adopt different analysis tools as we need them and incrementally improve the data collection and function over time.

**6.2 Open government data**

The US and UK have taken significant steps to make government data available. The US has made almost 160,000 data sets available online and the UK almost 27,000 data sets available online. Dataset available range widely from monthly house prices and income distribution, to local government and infrastructure, from biodiversity and food production, to geospatial data and environmental. The objectives are to increase transparency into government, to create

---


13 See the US government web site [http://www.data.gov/metrics](http://www.data.gov/metrics)

14 See the UK government website [http://data.gov.uk/data/search](http://data.gov.uk/data/search)
opportunities for businesses and to improve the quality of the natural environment by using data for monitoring.

An open data policy also has the potential to dramatically impact the performance of government itself as agencies are substantial producers and consumers of an enormous variety of data, great deal of which is relevant to more than once agency. As a trivial example, the Department of Education knows where the schools are and needs to know where the students will come from in future. The Department of Human Services knows where the students are coming from and needs to know where the schools are. Both departments collect data on both issues.

Increasingly, Australian State and Federal Governments have promoted a policy position that makes more government data sets available and which leverages the internet to engage more effectively with citizens, improve interaction between government agencies and with other sectors.

Despite repeated calls from both Federal and State Ministers, Australia appears to be relatively slow to open up data sets. The Australian Federal government website offers approximately 7,000 datasets\textsuperscript{15}. The New South Wales government currently has less than 400\textsuperscript{16}. There are many reasons for this apparent reluctance. Data custodians are concerned about the unintended consequences of release of data, in particular, the impact on individual privacy. Additionally, the cost of making data available, and concerns about data quality (and the conclusions drawn from incomplete or inaccurate data) are also often cited reasons for not releasing data.

6.3 Sharing data within government

Beyond concerns about unintended consequences of sharing anonymised data, a major challenge for government is existing regulation which prevents agencies from sharing data.

The Department of Human Services collects and uses a wide range of personal, health and sensitive information in order to deliver services, payments and carry out their mandated functions. Information is also used for planning, funding, administering, monitoring, evaluating, integrating and improving services and functions. Where appropriate and practicable, the DHS removes identifying details from information used for these purposes.

DHS collects information directly from individuals or third parties (family members, other government agencies or private organisations) where necessary. When the DHS collects personal information, it must be reasonably necessary for, or directly related to, the payments or services that they deliver. The type of personal information collected will depend on the kind of payment or service that being delivered, and the legislation under which the payment or service is provided. The collection of personal information may be specifically required or authorised by law, operating under 14 major Commonwealth Acts ranging from the Australian Hearing Services Act 1991 (Cth) to the Medical Indemnity Act 2002 (Cth).

The DHS privacy policy\textsuperscript{17} states that:

\textsuperscript{15} See the Australian Federal government website http://www.data.gov.au/dataset
\textsuperscript{16} See the NSW government website http://data.nsw.gov.au/data/dataset
Your personal information will not be shared across our service areas unless you have given your consent, or the sharing of your information in this way is authorised or required by or under law or the use is otherwise permitted by the Australian Privacy Principles. \footnote{Available online http://www.oaic.gov.au/privacy/privacy-resources/privacy-fact-sheets/other/privacy-fact-sheet-17-australian-privacy-principles}

We conduct data matching activities in accordance with legislation and legally binding guidelines issued by the Office of the Australian Information Commissioner. We comply with a Programme Protocol for each of our data matching programmes in accordance with guidelines issued by the Office of the Australian Information Commissioner.

In practice this means that services which could be created and be extremely valuable to citizens are currently not possible due to legislative limitations on data sharing. The example given in Section 3.4 of identifying ‘trigger points’ would be facilitated by the ability to share personal data much more widely within government.

The focus on privacy is understandable given the sensitivity of the information. The challenge to improving government productivity and improving service quality for citizens is navigating the myriad of relevant legislation required to effectively share data. Part of the role of the DTO is to identify ICT and data solutions which implement sharing and comply with this legislation.

**Box 20: Australia 3.0**

The Australia 3.0 \footnote{See website for further information www.australia30.org} series of workshops recently reviewed data challenges in the health sector. This sector, including provision of aged care services, overtook the retail sector in 2011 as the largest employer in Australia and yet it continues to struggle with growing challenges in health service delivery. Despite the estimated spend of $147 billion in 2012-13 \footnote{AIHW, ‘Health expenditure Australia 2012-13’, September 2014. Available online http://www.aihw.gov.au/publication-detail/?id=60129548871}, the sector faces long term challenges including the changing case mix driven by Australia’s ageing population, and substantial increases in levels of chronic disease.

Three significant developments over the past five decades make this a major public policy and economic challenge: firstly, the developments within our healthcare system to address all types of diseases with interventions and pharmaceutical support to reduce their impact on quality and life expectancy; secondly, lifestyle changes dominated by the rise in chronic health conditions such as diabetes, cardiac conditions and their resultant negative impacts on health and workforce productivity, and finally, increased life expectancy of an aging population with more than one chronic condition.

Bridging the gap between innovation and adoption will be critical to addressing this growing challenge. Health care is such a complex system of systems that modelling and simulation – the flow of patients through the health care system, changes to payment systems, the introduction of new technologies or treatment procedures, and construction of new hospitals – are critical components of future planning, similar in many ways to clinical trials. Datasets and knowledge are the foundation...
to ensure that we are focused on the most critical and valuable interventions to transform delivery.

A ‘Citizen centric’, integrated care system is the current focus of the countries with a commonwealth system. This requires the joining of available data sets to ensure that trial and error is not the basis of change. The data sets which would drive the greatest change include:

- Health Workforce data
- Medicare & PBS data
- Public and private hospitals data
- Health insurance claims data
- Disability data
- Mental health data
- Residential aged care data
- Community aged care data
- Data on key health and aged care

These data sets are the most valuable in addressing the challenges in health in the sense that they would allow modelling and simulation trials to test new forms of service delivery, including the impact of telehealth.

6.4 Broader privacy considerations in an Australian context

In Australia, the use of ‘personal information’ is primarily regulated by the Federal Government’s Privacy Act 1988 (Cth)\(^{21}\) (the Act). The Act includes the Australian Privacy Principles (APP) that apply to the handling of personal information by Australian Government agencies and many private sector organisations. The ‘principles’ are in fact legally binding\(^{22}\).

The Act includes credit reporting provisions that apply to the handling of credit-related personal information that credit providers are permitted to disclose to credit reporting bodies for inclusion on individuals’ credit reports. The Act regulates the collection, use, storage, security, disclosure, and disposal of individuals’ tax file numbers. The Act is also intended to implement Australia’s privacy obligations under the International Covenant on Civil and Political Rights\(^ {23}\) and to give effect to the OECD Guidelines on the Protection of Privacy and Cross Border Flows of Personal Data. The Act governs the private sector, including corporations and other businesses.

There are criminal penalties under the Privacy Act for unauthorised access to and disclosure of credit reporting personally identifiable information. Criminal sanctions also apply to the unauthorised disclosure of PII during an emergency or disaster situation. The Australian federal police would investigate such offences.

---


Box 21: Optus

In early 2015, telecommunications provider Optus was fined for breaches of personal data which affected more than 300,000 customers\textsuperscript{24}. The data breaches fueled the current debate about the Federal Government’s recently announced data retention legislation\textsuperscript{25}.

Australia’s largest mobile operator Telstra has confirmed it will to secure the non-content data it is now legally required to store on customers for two years.

The data retention bill includes a provision for mandatory data breach notification, which will require telecommunications operators and internet service providers to notify the Privacy Commissioner in the event of any breach. The notification scheme is scheduled to be introduce before the end of the 2015.

7. Conclusion – Australia has a productivity challenge

Increasing productivity in the digital services sector has become the new benchmark for international performance and Australia has fallen behind many of our trading partners. Productivity growth in the digital (services) economy means taking advantage of advances in ICT, as well as increasing use of data available from government, industry and citizens.

In the rapidly growing number of ‘as a platform’ companies operating squarely in the digital economy, data sets have become a major driver of output, and so productivity, greatly outpacing the importance of land, labour or capital. In the digital economy, data deserves to be considered a primary factor of production in its own right.

Pervasive use of data means organisations can continually experiment, simulate and analyse outcomes to better understand their own business, improve operational performance, guide investment decisions and improve customisation of services. This can help reduce the variability of outcomes while improving financial and product performance.

The finance and insurance services sector is the largest sector of the Australian economy, and whilst the sector has historically been a rapid adopter of ICT, the sector has performed modestly from a productivity growth perspective. Changing customer expectations and new waves of technology innovation are challenging traditional providers. Small, nimble companies are attacking parts of the traditional value chain – creating new services by harnessing a wide range of large, constantly evolving and highly personalised data sets: bigger, better data.

In finance and insurance, new technology and data driven services are offering immediate benefits to a growing class of consumers willing to move from traditional providers and from traditional financial products. It is important for the productivity of the sector to allow new technology and new business models to flourish. It also remains important for regulators to keep pace with new technologies and develop agile frameworks which are supportive of the introduction of new services,


Available online http://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/Bills_Search_Results/Result?bId=r5375
whilst at the same time ensuring the protection of consumers (including protecting consumer personal information).

Bigger, better data will increasingly become synonymous with open data, allowing consumers and citizens to make decisions on a much more equal footing with larger institutions that once had a monopoly on critical information. Australia is well behind many other developed countries in the number, quality and richness of data sets which are made publically available.

In areas as complex as health and human services, access to data held by governments will be critical to understanding and addressing some of the greatest challenges facing Australia. The perceived and actual protection of private individual data remains the great challenge in opening up these data sets. Policy reform is required to clarify under what conditions governments can make public data sets (in a form which strips out personal information) accessible to researchers, private users and companies who can generate new value from them.

Leadership will be required, and policy reform may be required, to encourage the cultural change needed for government agencies to move from reactive disclosure of data to a more open, proactive data release approach.

The prize is the opportunity to create benefit for Australian industry, benefit for government, greater decision making transparency for the citizens of Australia, while still protecting the rights and the sensitive, personal information associated with each of us as individuals.

7.1 Linking ‘bigger and better data’ to the Financial System Inquiry’s recommendations

The Financial System Inquiry recommended that the Federal Government:

> Review the costs and benefits of increasing access to and improving the use of data, taking into account community concerns about appropriate privacy protections.

This section makes further recommendations for access and use of data

**Recommendation 1**: Regulatory clarification – Australia should have the goal to be in the top countries in the OECD in terms of number, type and quality of data shared.

The benefits of sharing data in a way which preserves individual privacy have been a consistent theme throughout this paper. The government sector is one of the greatest curators and consumers of data in Australia. The regulatory complexity outlined in Section 6 highlights the challenges associated with greater sharing of government data. It is far too easy to read ‘not allowed’ into existing regulations at one or more levels and so effectively prevent opening up of government data. Clarifying regulations associated with the release and use of government data will help encourage different government agencies to open up and share data. Efforts have been made at both Federal and State levels, however adoption of a more proactively open data policy which asks ‘if not, why not?’ framed within the scope of existing legislation would go a long way to providing the necessary clarity many agencies need to open up data sets.
Organisations such as the Open Data Institute (UK)\(^{26}\), the Sunshine Foundation (US)\(^{27}\) and Privacy By Design (Canada)\(^{28}\) provide principles for proactive disclosure which also seek to maximise individual privacy. In this area, Australia has an opportunity to adopt the good work done by these international organisations.

**Recommendation 2:** Developing a regulatory framework which supports new technology – encourage regulators to work with Industry to understand the risks and implications of new technology as it becomes a more significant force in the market.

It is difficult to foresee the long term impact of much of the new technology being released into the market. Technology is evolving so rapidly that it is also difficult to categorise according to traditional regulatory frameworks. What remains true is the need for government to provide regulation to mandate fair dealing, to help consumers to understand risk, and to minimise large adverse outcomes. As discussed in Section 5, peer-to-peer lending, crowd sourced funding and other new, technology driven business models are examples where regulators should work closely with the developing industry players to support the innovation whilst protecting the public from unscrupulous players or adverse risk. Almost uniquely, regulators have the opportunity to work collaboratively with industry and consumer groups to evolve codes of practice and regulation which allow innovation to flourish at the same time as protecting consumers. This requires new, agile thinking from regulators which is appropriate given the agile, discovery-directed nature of much of the new technology being developed. In each of the areas of described in this paper – peer-to-peer lending, crowd funding, robo-advice – regulators have the opportunity to learn from international experience and iteratively develop codes of practice and regulation.

**Recommendation 3:** Research on data sharing – a framework should be developed which supports anonymization of data and in turn facilitates sharing.

One of the major concerns of technology driven business is the potential to identify individuals and gain access to personal or sensitive information by collating multiple data sources. This is a factor limiting uptake of online services, particularly with government as described in Section 6. Focussed research is needed to develop robust methods of sharing data in a way which maintains privacy within an Australian regulatory context. The areas which have the greatest potential to drive productivity in Australia are also the areas which require access to the most sensitive and personal data sets – health, superannuation, human services, and education. A focussed effort on mechanisms which allow data to be anonymised and shared with industry and the research community will open up many of the biggest challenges facing Australia to the academic scrutiny and industry led innovation.

\(^{26}\) See organisation website for more information [https://theodi.org/about](https://theodi.org/about)

\(^{27}\) See organisation website for more information including ‘proactive disclosure’ principles [http://sunlightfoundation.com/opendataguidelines/](http://sunlightfoundation.com/opendataguidelines/)

\(^{28}\) See organisation website for more information [https://www.privacybydesign.ca/](https://www.privacybydesign.ca/)
Recommendation 4: Research data – a framework should be developed to provide data sets to the Australian research community.

A related point to the Recommendation 3 is providing access to anonymised data to better understand large scale, complex problems. Throughout this paper, examples are given of the potential value which can be created from data sets generated by users and by governments. The most valuable data sets for researchers are however often the ones which are most sensitive. Whilst Recommendation 3 seeks a means of sharing data widely, there is value in sharing data with research institutes to explore issues of subtlety and complexity for non-commercial purposed. Providing a framework under which research organisations can gain access to anonymised versions of significant, national data sets is crucial to understanding some of the most significant national challenges including health, the implication of an aging population and national productivity.

Recommendation 5: Spatial data – should be brought under one standard and released as open source.

As outlined in Section 2.2, spatial data sets have some of the greatest potential for economic impact supporting use cases from improving transport and logistics to helping resolve property disputes. The number of use cases for spatial data has increased dramatically as more services come to include a spatial component. All data ‘happens’ somewhere and so the spatial component of data associated with services, vehicles, machines, buildings, infrastructure, livestock, weather and people brings added richness to new service creation. Just as we once debated the need to develop a single standard railway gauge, the multiple frameworks for recording spatial data should be harmonised under one national standard. Further, data from the various Australian state and federal data sets should be released as under open source licence. Australian spatial data is more likely to create value for Australia than any other data sets.

Recommendation 6: Data breaches – mandatory disclosure of breaches of data and cyber incursions.

The airline industry is exemplary in the rigorous investigation of airline incidents and the public disclosure of the outcomes. Partly as a consequence of the rigour and public scrutiny associated with these detailed investigations, and the associated modifications of technology and protocols within the airline industry, air travel is arguably the safest of all modes of travel for passengers.

The strong trend towards increasing digitisation of services described throughout this paper is premised on confidence in the security of the underlying technology and of the data used to drive those services. Building confidence in new technology and approaches comes from honest declarations of when things have gone wrong.

It can be argued that cyber incursion represents a different challenge to all other forms of attack or major incident. Distance is no longer a barrier, attacks may come from individuals or nation states, and the ‘attack’ surface (as a consequence of use of mobile devices and distributed computing systems) is constantly evolving making it difficult to completely lock down all points of entry to a system. Nonetheless, mandatory disclosures of data breaches and cyber incursions to a public who
are not concerned with the intricacies of technology, would strengthen trust in digital economy services and should apply to all industry sectors including government.

Concerns that disclosure of breaches put organisations at further risk can be addressed through appropriately timed release of detailed information to the public once systems have been made more robust.

Concerns raised about the impact on company valuation or shareholder flight can be evaluated against actual company disclosure information. Due to the continuous disclosure requirements in Australia, the ASX processes more than 130,000 company announcements each year, including those considered to be price sensitive so leading to temporary suspension of trade whilst announcements are processed. More than a decade of tick-by-tick (microsecond rate) trading records exist for the ASX showing market reactions to company announcements – including those involving cyber incursion or data breach. These records are available to be analysed and can be used to provide a historical understanding of the intraday and end of day price impact on a company announcing a cyber incursion or data breach on the ASX.

Recommendation 7: Developing a regulatory framework which supports victims of data breach – adopting the philosophy that it is better to plan how to deal with cases of violation of individual privacy rather than look for solutions after the fact, regulated actions plans and a framework for compensation should be developed. This complementary recommendation to that above would help build confidence in a world of increasingly digital services. Frameworks already exist for credit card and payments fraud, so the recommendation is to extend this to other forms of digital services.

Recommendation 8: An Accounting Standard for Data – As argued in Section 2, in the Digital Economy, data should be considered a primary factor of production. Unless the value of data can be estimated in an accounting sense, data will be undervalued as a factor of production in the Digital Economy. With more than 100 years of significant international contribution through Standards Australia, Australia has an internationally respected reputation as an authority in practical and relevant standards development. The recommendation is to commission a detailed, focussed effort to develop a national accounting framework for Bigger Better Data which goes beyond treating data as a footnote in intellectual property accounting.
References


Forbes, 2014a. *How Target figured out a teen girl was pregnant before her father did.*


http://www.mckinsey.com/insights/globalization/global_flows_in_a_digital_age


http://www.mckinsey.com/~/media/mckinsey%20offices/singapore/2014%20digital%20banking%20in%20asia%20%20winning%20approaches%20in%20a%20new%20generation%20of%20financial%20services.ashx


http://www.publications.parliament.uk/pa/cm201415/cmselect/cmtreasy/204/20408.htm


Van Onselen, L., 2014. *It’s housing that’s killing productivity.*


http://mathworld.wolfram.com/BayesianAnalysis.html


About the Australian Centre for Financial Studies

The Australian Centre for Financial Studies (ACFS) seeks to enhance excellence in financial services by promoting a strong research network among industry, government and academic institutions. The Centre specialises in industry-oriented research and events, aiming to boost the global credentials of Australia’s financial sector and support Australia as an international centre for finance research, practice, and education.

The Centre is a consortium of Monash University, RMIT, and the professional body Finsia, with Associate University members the University of Melbourne, Griffith University and Victoria University. ACFS receives support from a number of corporate sponsors, and the Victorian Government.

www.australiancentre.com.au

ACFS Research Code of Conduct

ACFS conducts independent research and thought leadership into issues related to the financial sector. Contracted research is undertaken for industry research partners, or for the purpose of generating public debate about issues of importance to the sector and the wider community.

The strength of ACFS research is that it is well informed by industry practice and input, and is undertaken with a high level of expertise, by ACFS research staff and by drawing on our wide network of academic and practitioner collaborators.

We produce accessible and relevant research that is non-partisan, and non-sectoral in its perspective. While we welcome a diversity of views, the value of our work is underpinned by its independence.