



# **Costs and Benefits of Data Provision**

Report to the Australian National Data Service

By

**John Houghton**

September 2011

**Centre for Strategic Economic Studies  
Victoria University**

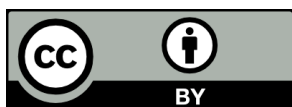
## **Acknowledgements**

The author would like to acknowledge the support of The Australian National Data Service (ANDS), and thank Dr. Greg Laughlin of ANDS for all his efforts during the project, and all those from the case study agencies who provided the necessary data for the study and generously gave their time for interviews and questions.

## **Disclaimer**

While every effort has been made to ensure its accuracy, neither Victoria University nor the Centre for Strategic Economic Studies make any representations or warranties (express or implied) as to the accuracy or completeness of the information contained in this report. Victoria University and the Centre for Strategic Economic Studies, their employees and agents accept no liability in negligence for the information (or the use of such information) provided in this report.

September 2011



This work is licensed under the Creative Commons Attribution 2.5 Australia License.

To view a copy of this license, visit <http://creativecommons.org/licenses/by/2.5/au/> or send a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.

## Summary of findings

Over the last decade there has been increasing awareness of the potential benefits of more open access to Public Sector Information (PSI) and the findings of publicly funded research. That awareness is based on economic principles and evidence, and it finds expression in policy at institutional, national and international levels.

Public Sector Information (PSI) policies seek to optimise innovation by making data available for use and re-use with minimal barriers in the form of cost or inconvenience. They place three responsibilities on publicly funded agencies: (i) to arrange stewardship and curation of their data; (ii) to make their data readily discoverable and available for use and re-use with minimal restrictions; and (iii) to forgo fees wherever practical.

This report presents case studies exploring the costs and benefits that PSI producing agencies and their users experience in making information freely available, and preliminary estimates of the wider economic impacts of open access to PSI. In doing so, it outlines a possibly method for cost-benefit analysis at the agency level and explores the data requirements for such an analysis – recognising that few agencies will have all of the data required.

There are many ways in which the provision of more open access to PSI can impact upon the costs faced by the government agency producers and the many existing and potential users of the information. This study focuses on three main elements:

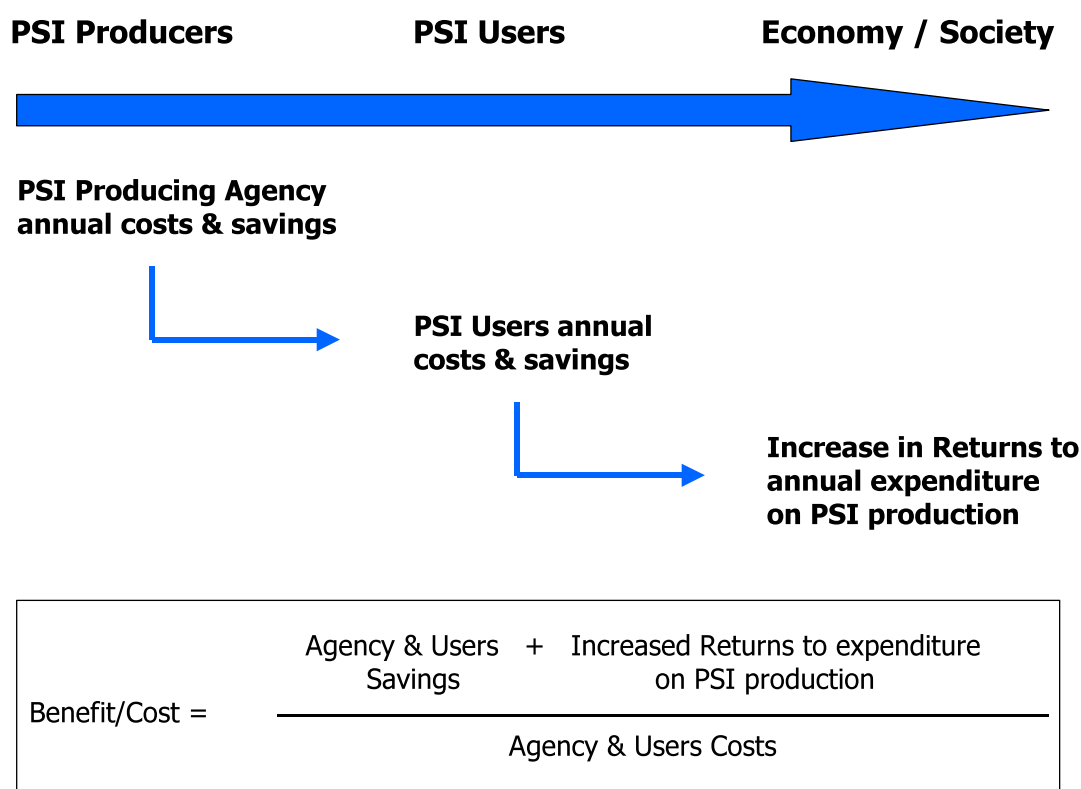
- The costs and cost savings experienced by PSI producing agencies involved in the provision of free and open access to information;
- The costs and cost savings experienced by the users of PSI in accessing, using and re-using the information; and
- The potential wider economic and social impacts of freely accessible PSI.

It is always more difficult to identify benefits than costs. Benefits may accrue in a variety of ways, including cost savings, efficiency gains, and new opportunities to create value through doing things in new ways and doing new things. These are, successively, more difficult to quantify: not least because they often emerge over time and can only be realised in the future. An obvious approach is to begin with the most direct and directly measurable benefits, namely agency and user cost savings. Wider benefits are more difficult, and in some cases impossible, to measure. In this study, we explore impacts on consumer welfare and attempt to estimate the impacts of increased access and use, as measured by increased downloads, on returns to expenditure on data production.

While there are some one-off costs involved in the change to open access, most are recurring annual costs (*e.g.* agency IT and hosting costs, revenues foregone, etc.). Hence, both the agency and user costs that are modelled are annual costs, and the cost savings annual savings. In terms of the wider benefits of open access to PSI, returns to investment in data production are recurring annual returns, lagged and discounted over the useful life of the data – using a perpetual inventory method. Consequently, the cost-benefit comparisons presented in this study

include annual agency and user costs and cost savings as well as the wider benefits arising from increased returns to annual expenditure on data production (Figure 1). They compare the costs and benefits at the time of the transition to open access (*i.e.* at the prices and levels of activity of the time).

Figure 1 A framework for estimating cost-benefit



Source: Author's analysis.

It is clear from the case studies presented that even the subset of benefits that can be measured outweigh the costs of making PSI more freely and openly available. It is also clear that it is not simply about access prices, but also about the transaction costs involved. Standardised and unrestrictive licensing, such as Creative Commons, and data standards are crucial in enabling access that is truly open (*i.e.* free, immediate and unrestricted).

For example, we find that the net cost to the Australian Bureau of Statistics (ABS) of making publications and statistics freely available online and adopting Creative Commons licensing was likely to have been around \$3.5 million per annum at 2005-06 prices and levels of activity, but the immediate cost savings for users were likely to have been around \$5 million per annum. The wider impacts in terms of additional use and uses bring substantial additional returns, with our estimates suggesting overall costs associated with free online access to ABS publications and data online and unrestrictive standard licensing of around \$4.6 million per annum and measurable annualised benefits of perhaps \$25 million (*i.e.* more than five times the costs).

While data are more limited, there appears to have been an even more compelling case for making fundamental geospatial data freely available. Of course, the relative cost-benefits apply to the form of PSI involved and do not reflect in any way on the performance of the producing agencies. Some forms of PSI underpin major industries and contribute to their growth and prosperity. Other forms of PSI may have an important influence on policy decisions, but the economic impacts may be more limited and difficult to trace.

The publications and data arising from publicly funded research differ somewhat from other forms of PSI. Consequently, it is difficult to draw direct lessons for the research sector from the case studies explored in this report. Nevertheless, it is clear that many of the same issues arise when attempting to measure the value of the information and/or the costs and benefits associated with providing open access to it.

The evidence from previous studies suggests that individual cases vary greatly, making generalisation extremely difficult. Perhaps, what could more usefully be generalised are the methods of analysis. For example, it would be useful to combine the frameworks and models into a tool that could be applied in assessing the costs and benefits of research data curation and sharing, and to further develop the framework for estimating cost-benefits outlined in this study to produce a tool tailored to the analysis of the costs and benefits of providing open access to PSI. These tools might consist of a template for data collection, a draft questionnaire outlining the questions needed to elicit the necessary information, and a simple spreadsheet-based online model that people could use to perform a cost-benefit analysis. The models should include all possible quantifiable costs and benefits, but must also include qualitative issues to help to prioritise data preservation, access and curation projects (*e.g.* incorporate a balanced scorecard approach to weighing the more intangible benefits).

What this study demonstrates is that the direct and measurable benefits of making PSI available freely and without restrictions on use typically outweigh the costs. When one adds the longer-term benefits that we cannot fully measure, and may not even foresee, the case for open access appears to be strong.

---

# Contents

<b>SUMMARY OF FINDINGS .....</b>	<b>III</b>
<b>1 BACKGROUND AND CONTEXT .....</b>	<b>1</b>
1.1 ECONOMIC PRINCIPLES .....	1
1.2 ECONOMIC EVIDENCE .....	2
1.3 POLICY RESPONSES .....	3
1.4 MEASURING VALUE, COSTS AND BENEFITS .....	5
<b>2 A FRAMEWORK FOR ESTIMATING COST-BENEFIT .....</b>	<b>7</b>
2.1 AGENCY COSTS AND COST SAVINGS .....	8
2.2 USER COSTS AND COST SAVINGS .....	10
2.3 EFFICIENCY AND PRODUCTIVITY IMPACTS .....	10
2.4 WIDER ECONOMIC IMPACTS AND BENEFITS .....	11
2.5 A COST-BENEFIT MODEL .....	12
2.6 GUIDE TO DATA REQUIREMENTS .....	13
<b>3 PUBLIC SECTOR INFORMATION CASE STUDIES .....</b>	<b>16</b>
3.1 NATIONAL STATISTICS (AUSTRALIAN BUREAU OF STATISTICS) .....	18
3.1.1 <i>Agency costs and benefits</i> .....	19
3.1.2 <i>User costs and benefits</i> .....	22
3.1.3 <i>Indicators of use (web statistics)</i> .....	23
3.1.4 <i>Wider impacts of use</i> .....	28
3.1.5 <i>Summary of impacts</i> .....	31
3.2 SPATIAL DATA (OFFICE OF SPATIAL DATA MANAGEMENT & GEOSCIENCE AUSTRALIA) .....	32
3.2.1 <i>Agency and user costs and benefits</i> .....	33
3.2.2 <i>Indicators of use (web statistics)</i> .....	34
3.2.3 <i>Wider impacts of use</i> .....	37
3.2.4 <i>Summary of impacts</i> .....	39
3.3 HYDROLOGICAL DATA (NATIONAL WATER COMMISSION & BUREAU OF METEOROLOGY) .....	40
3.3.1 <i>Agency and user costs and benefits</i> .....	41
3.3.2 <i>Indicators of use (web statistics)</i> .....	42
3.3.3 <i>Wider impacts of use</i> .....	44
<b>4 WIDER IMPACTS OF OPEN ACCESS TO PSI .....</b>	<b>46</b>
4.1 REPORTED IMPACTS OF SPATIAL DATA IN AUSTRALIA .....	46
4.2 IMPACTS OF PSI MORE BROADLY .....	48
<b>5 LESSONS FOR THE RESEARCH SECTOR .....</b>	<b>49</b>
5.1 COST-BENEFIT ANALYSES OF RESEARCH DATA CURATION AND SHARING .....	50
5.2 LESSONS FOR THE RESEARCH SECTOR AND NEXT STEPS .....	52
<b>ENDNOTES AND REFERENCES .....</b>	<b>54</b>

# Figures and Tables

Figure 1	A framework for estimating cost-benefit .....	7
Figure 2	PSI and public content domains with examples .....	17
Figure 3	ABS moves towards open access .....	18
Figure 4	Estimated annual savings from FPOW/FSOW (Circa 2006-07).....	19
Figure 5	Estimated ongoing annual loss in revenue due to FPOW/FSOW (Circa 2006-07) .....	20
Figure 6	ABS@ and Ausstats downloads .....	24
Figure 7	ABS website usage, 2003-04 through 2009-10 (re-based).....	26
Figure 8	Downloads per page view, page published and release, 2003-04 through 2009-10 .....	27
Figure 9	Scheduled spatial datasets delivered by government agencies, 2001-02 to 2005-06.....	36
Figure 10	Total page views or ‘hits’ reported by ASD, 1999-2010 .....	37
Figure 11	Data use and re-use from the Victorian Water Resources Data Warehouse .....	43
Figure 12	The KRDS2 benefits taxonomy .....	51
Table 1	Activity cost impacts for producers and users.....	9
Table 2	Estimated net position for ABS from FPOW/FSOW and CC licensing (per annum) ...	22
Table 3	Estimated annual user cost savings from free online and standardised CC licensing....	23
Table 4	ABS website usage, 2003-04 through 2008-09 .....	25
Table 5	ABS website usage, 2003-04 through 2009-10 (re-based).....	25
Table 6	Downloads per page view, page published and release, 2003-04 through 2009-10 .....	27
Table 7	Estimates of consumer surplus from ABS free online .....	29
Table 8	Estimates of increase in returns to annual spending from ABS free online.....	30
Table 9	Summary of ABS case study annual costs and benefits (circa 2005-06).....	31
Table 10	Spatial datasets delivered by Australian government agencies, 2001-02 to 2005-06 ....	35
Table 11	Australian government revenue from sales of spatial data, 2001-02 to 2005-06 .....	36
Table 12	Estimates of annual consumer surplus and increase in returns to annual spending .....	38
Table 13	Summary of spatial data case study annual costs and benefits (circa 2005-06).....	39
Table 14	Data use and re-use from the Victorian Water Resources Data Warehouse .....	44
Box 1	Maximising the value of PSI through enabling maximum use .....	6
Box 2	Environmental impact assessments in Europe .....	12
Box 3	Method: Welfare impacts and returns to expenditure.....	28
Box 4	The Victorian Water Resources Data Warehouse.....	42
Box 5	Australia’s Water Market.....	45
Box 6	Benefits of curating and sharing publicly funded research data .....	49





# 1 Background and context

Over the last decade there has been increasing awareness of the potential benefits of more open access to Public Sector Information (PSI), research publications (e.g. journal articles) and research data. That awareness is based on economic principles and evidence, and it finds expression in policy at institutional, national and international levels. In this section we briefly explore the economic principles, evidence and policy responses, before turning attention to the issue of measuring the value of enhanced access.

## 1.1 Economic principles

Information has public good characteristics (*i.e.* being non-rivalrous and non-excludable), as one person's consumption of a piece of information does not prevent others from consuming it and it is difficult to prevent information spreading to others. While information can be made more or less excludable through intellectual property rights, such as copyright, it is still difficult to stop people sharing information. In general, the private sector will tend to under-produce such goods as it is difficult to realise the full value. It is this that justifies public sector supply of information (Nilsen 2007; 2009).<sup>1</sup> Indeed, Stiglitz *et al.* (2000) concluded that the theoretical underpinnings of the private versus public trade-off shifts as the economy moves toward a digital one, with a larger public role in the digital economy.<sup>2</sup>

Economists argue that pricing above the marginal cost of dissemination is inefficient because it results in a deadweight loss and eliminates some of the consumer surplus – some people will be prevented from enjoying the benefit of the good even though their consumption of the good would come at little or no marginal cost to the producer. The marginal cost of online dissemination is close to zero. Cost recovery through such mechanisms as user fees is never welfare enhancing (Nilsen 2007; 2009).<sup>3</sup> For most PSI, marginal cost pricing will be the preferable option (Pollock 2009, p40).<sup>4</sup>

Referring to Statistics Canada's shift from cost recovery to free online access, Nilsen (2009, p43) concluded that:

*While one can predict that the agency's sales and licensing revenues will decrease, it is also true that usage and reuse will increase. Increased usage would have positive externalities in terms of information dissemination and the uses to which people will put the data, and it will also have a positive economic impact for the country as a whole, ultimately leading to increased tax revenues that are generated by the use and reuse of PSI.*

*At the same time, the agency's transaction and opportunity costs will decrease. A great deal of money and time is currently spent determining prices, negotiating and administering licenses, and monitoring where sales revenues are coming from, who is paying, and who has not paid. Furthermore, much time and money is spent negotiating with other government departments over cost recovery charges for information, which generates a great deal of wasted transaction costs and produces no new revenues for*

*the government as a whole. That money could be put to other uses. I conclude that the decreases in transaction and opportunity cost will more than compensate for the revenue decrease; meanwhile, there will be a positive economic impact for the country from increased use of the data and the tax revenue increases resulting from private sector reuse. Overall this outcome will result in increased economic efficiency and a greater net social benefit.*<sup>5</sup>

Some public sector agencies retain copyright on the information they produce to enable them to control access and the conditions of use, so that they can impose a toll on that access and use and prevent uses that would result in the loss of control. However, economists have concluded that government or “Crown” copyright has social costs and a negative economic impact (Nilsen 2009).<sup>6</sup> Moreover, if the price is set to the marginal cost of online dissemination (zero), there is no purpose to be served by retaining copyright. Indeed, as the costs of disseminating and accessing information have declined, the transaction costs associated with charging for access to information have come to constitute a major barrier to access in themselves (Quiggin 2010).<sup>7</sup> Studies of transaction costs relating to information suggest that they are significant (Poplin [nee Krek] 2010).<sup>8</sup> Hence the efficient economic solution for the dissemination of PSI is likely to be free *libre* and free *gratis* (*i.e.* making it freely available online *and* using unrestrictive licensing, such as Creative Commons).

## 1.2 Economic evidence

Evidence suggests that PSI plays an increasingly important role in knowledge-based economies. A number of studies have attempted to put a value on PSI, but they have used a variety of approaches and produced a variety of answers.<sup>9</sup>

By generalising from case studies and scaling up, PIRA (2000) estimated investment value of PSI (*i.e.* what governments invest in the acquisition of PSI) and economic value of PSI (*i.e.* the national income attributable to activities built on the exploitation of PSI) in the European Union, putting the former at around EUR 9.5 billion per annum in 1999 and the latter at around EUR 68 billion (equivalent to approximately 1.4% of EU GDP – a seven-fold return on investment).<sup>10</sup> By comparison, they put PSI investment in the United States at EUR 19 billion per annum and economic value at EUR 750 billion, suggesting that the EU could reach US levels with more open access regimes, but would only need to double the value of PSI for governments to recoup the lost revenues from PSI sales in increased tax receipts. The PIRA report was popularised by Peter Weiss (2002) in the widely cited “*Borders in Cyber-Space*” report.<sup>11</sup>

Employing a large survey of PSI producers and users, Dekkers *et al.* (2006) sought to estimate the size of the PSI market in Europe (*i.e.* the MEPSIR study). Based on the estimates of re-users they put the overall market for PSI in the EU plus Norway at around EUR 27 billion (approximately 0.25% of aggregated GDP).<sup>12</sup> This is a much lower number than suggested by the PIRA study, despite it being market size rather than value added and coming five years latter. Making some adjustments with the benefit of hindsight, te Velde (2009) suggested that the value might drop further from EUR 27 to EUR 5 billion or even EUR 3 billion<sup>13</sup> – only around 5% of the PIRA study estimate, and less than PIRA’s estimate of investment value.

DotEcon (2006) produced a report for the UK Office of Fair Trading (OFT) in which they developed a bottom-up approach to estimating the economic value of PSI products and services in the UK, seeing the net economic value of PSI as the willingness to pay for PSI minus the cost of supplying it (essentially, the net consumer surplus). They also looked at the costs (detriment) of barriers to use, including: unduly high prices, distortion of downstream competition, and failure to exploit PSI.<sup>14</sup> The results indicated that the net value of PSI in the UK was around GBP 590 million per annum in 2005. The costs of the three types of detriment were estimated to be GBP 20 million from high pricing, GBP 140 million from restriction of downstream competition, and GBP 360 million from failure to exploit PSI – suggesting that the value of PSI could be doubled by resolving the problems identified.<sup>15</sup>

Looking at spatial information in Australia, ACIL Tasman (2008) estimated that industry revenue in 2006-07 could have been of the order of \$1.37 billion and industry gross value added around \$682 million. Using a General Equilibrium (GE) modelling approach, they concluded that the economic footprint of the spatial information industry was larger. Accumulated impacts:

- Contributed to a cumulative gain of between \$6.43 billion and \$12.57 billion in Gross Domestic Product (GDP) – equivalent to 0.6% and 1.2% of GDP, respectively;
- Increased household consumption by between \$3.57 billion and \$6.87 billion on a cumulative basis;
- Increased investment by between \$1.73 billion and \$3.69 billion on a cumulative basis;
- Had a positive impact on the trade balance – as exports were between \$1.26 billion and \$2.30 billion higher than they would otherwise have been, and imports were between \$1.18 billion and \$2.23 billion higher than they would otherwise have been; and
- Real wages were between 0.60% and 1.12% higher than they would otherwise have been.<sup>16</sup>

What these example studies show is that different approaches to estimating the value of PSI produce very different answers, but the common feature of these and many other studies is that the economic and social value can be high, often far outweighing the costs of collection and dissemination.

### 1.3 Policy responses

Defining PSI as “including information products and services, generated, created, collected, processed, preserved, maintained, disseminated, or funded by or for the Government or public institution,” OECD (2005) explored the economic issues and provided an overview of developments within OECD countries at that time. They concluded that access, pricing and licensing policies make a substantial difference in the levels of access to, use and re-use of PSI.<sup>17</sup>

The *OECD Recommendation on Public Sector Information*, adopted by the OECD Council on 30 April 2008, provided policy guidelines designed to improve access and increase use of PSI

through greater transparency, enhanced competition and more competitive pricing.<sup>18</sup> *Inter alia*, it recommended:

- Maximising the availability of PSI for use and re-use based upon the presumption of openness as the default rule;
- Strengthening awareness of what PSI is available;
- Ensuring methodical data collection and curation practices to enhance quality and reliability; and
- Maximising the integrity and availability of PSI through the use of best practices in information management.

These principles were derived from existing best practice and became the basis for further developments in PSI access regimes in OECD countries and elsewhere around the world.

Similar principles were reflected in the recommendations of the Australian Government 2.0 Taskforce, which concluded that government should make PSI open, accessible and reusable, suggesting that by default PSI should be: free, based on open standards, easily discoverable, understandable, machine-readable, and freely reusable and transformable.<sup>19</sup> In turn, these principles were embodied in last year's Declaration of Open Government and given force in the passage of legislation reforming the Freedom of Information (FOI) Act and establishing the Office of the Australian Information Commissioner, the Government's response to the Government 2.0 Taskforce report and its response to the *Ahead of the Game: Blueprint for the Reform of Australian Government Administration* report.<sup>20</sup>

Similar principles underpin policies on research publications and research data. Around the world, more and more research institutions and funders are introducing Open Access (OA) policies that request and sometimes demand that research publications and, increasingly, research data be made available in OA journals and/or repositories. Examples include major research funders, such as the US National Institutes of Health (NIH) and the UK's Wellcome Trust, and such institutions as Massachusetts Institute of Technology (MIT) and CERN. In Australia, the Australian Research Council (ARC) and National Health and Medical Research Council (NHMRC) both mandate Open Access to the results of their funded research, and a number of universities mandate Open Access to research performed at the institution (including one of the world's earliest adopters of an OA mandate, Queensland University of Technology).<sup>21</sup> Some countries are going one step further with what amounts to a national OA mandate (e.g. Brazil and the proposed US *Federal Research Public Access Act*).

In all cases, the underlying principles are that PSI, research publications and data should be:

- Freely available at no cost, or no more than the marginal cost of distribution;
- Easily discoverable and readily accessible; and
- Open to unrestricted use and re-use on conditions that are readily discoverable and understandable.<sup>22</sup>

As Eves (2009) put it: if it can't be spidered or indexed, it doesn't exist; if it isn't available in open and machine readable format, it can't engage; and if a legal framework doesn't allow it to be re-purposed, it doesn't empower.<sup>23</sup>

## 1.4 Measuring value, costs and benefits

As noted, there have been a number of approaches to measuring the value of PSI and estimating the potential benefits of access and re-use. These have included top-down econometric modelling, extrapolations based on surveys of PSI producers and users scaled to national or regional markets, estimates based on agency costs and consumers' willingness to pay, and estimates of elasticities and multipliers.

PIRA (2000) combined measures of the investment cost (*i.e.* the amount spent on the collection/generation of the PSI) and expenditure on PSI by users and re-users, then: for final users, estimated value as expenditure on PSI or, where the PSI was freely available, as the investment cost of its collection/generation; and for intermediaries who add value to the data and re-sell, ascribed a proportion of their value added to the PSI used. A simplifying assumption was that the value of the information accrued in the year that is collected/generated. This is reasonable for some PSI, such as weather forecasts, less so for mapping information and entirely inappropriate for research publications and data, which may take years to realise value in application and commercialisation (and were excluded from the PIRA study). Aside from the many difficulties in estimation and attribution, a potential problem with the PIRA approach is that it may overestimate the value of PSI because it does not account for the possibility of the use of alternatives.

In the MEPSIR study of Dekkers *et al.* (2006), demand and economic performance were measured in an extensive survey by directly asking both public content holders and re-users for key economic data, such as total turnover against turnover related to PSI, total number of staff against the number of staff dedicated to handling PSI, and estimates of domestic market for a particular type of PSI. The European PSI market value was then estimated from the average revenues multiplied by the average number of re-users per PSI domain, minus the cost of PSI collection/generation. For country estimates, this was distributed according to GDP. With the benefit of hindsight, te Velde (2009) suggested that the estimates were high, perhaps 5 to 9 times too high.

DotEcon (2006) adopted a bottom-up approach to estimating the economic value of PSI products and services in the UK, seeing the net economic value of PSI as the willingness to pay for PSI minus the cost of supplying it. Using a survey and published sources, this was estimated from: (i) the net consumer surplus from PSI (*i.e.* the amount that customers might be prepared to pay over and above what they do pay to have access); and (ii) the total producer surplus that arises from the provision of PSI (*i.e.* the extent to which revenues exceed the costs of supplying the product or service). Adding these two estimates gave the net economic value of PSI. While much less subject to over-estimating the value of PSI, a potential weakness of this approach lay in estimating price elasticities of demand, especially where the PSI was supplied free of charge.

ACIL Tasman (2008) used a value-added approach based on General Equilibrium (GE) modelling, with input data derived from case studies, which were used as a guide to estimating

the direct impacts of spatial information on selected sectors in the Australian economy, and a GE approach to modelling economy-wide impacts. A potential issue with this approach is how to scale from case studies to sector-wide impacts (*i.e.* understanding the relationship between the cases and the sector).

### Box 1      Maximising the value of PSI through enabling maximum use

The Earth's atmosphere, oceans and landscapes are changing rapidly, with human activities being a major driver. Monitoring and modelling these changes are critical to enabling governments, civil society and the private sector to make informed decisions about climate, energy, food security, natural hazard, health and other challenges.

Many empirical studies of the dissemination of public sector data and information (as well as publicly funded data in academia) have repeatedly demonstrated the value of open access to such information online. The lack of restrictions on their re-use provides much greater economic and social returns than proprietary dissemination systems with access costs and re-use restrictions. Real life cases studies mirror some of the conclusions of the empirical studies:

- CBERS (China Brazil Earth Resources Satellite) removal of imagery charges resulted in increased access from 1,000 images per year to 10,000 images per month with more than 10,000 new users registered in the first year. Ninety-eight per cent of users surveyed agreed with the policy of open data access and reported the creation of many new jobs, the creation of new SMEs and improved research and teaching;
- The US Geological Survey achieved similarly impressive results following removal of Landsat's charges for Internet users, which resulted in more Landsat data (more than 1.1 million images) being processed and distributed in FY 2009 than in the whole 38 year mission history combined;
- Conversely, the ASTER Global Digital Elevation Model (GDEM) saw a reversal in fortunes following the introduction of charges in January 2006, with a substantial reduction in data distribution. A reversal in policy to no charge in June 2009 again provided a clear indication of dramatically increased usage of ASTER data worldwide (with over 6.5 million tiles distributed in FY 2009);
- The Argo program's public data access policy has enabled the provision of near real-time access to the first continuous global monitoring array of ocean temperature, salinity and velocity for use in oceanographic and climate forecasting models; and
- GBIF's aims to make scientific biodiversity data the common property of everyone has resulted in the publication online of nearly 200 million biodiversity records from 300 data publishers in 54 countries.

Improving data access and sharing systems significantly increases data value by reducing the cost and re-use restrictions for users.

Source: GEO (2010) *GEOSS Data Sharing Action Plan*. GEO-VII Plenary.

While all approaches have their limitations, such studies suggest that bottom-up and top-down approaches, macroeconomic and microeconomic approaches are all feasible. Perhaps the best approach is to try to combine them, and estimate the value of PSI and of making PSI more openly accessible by means of triangulation.

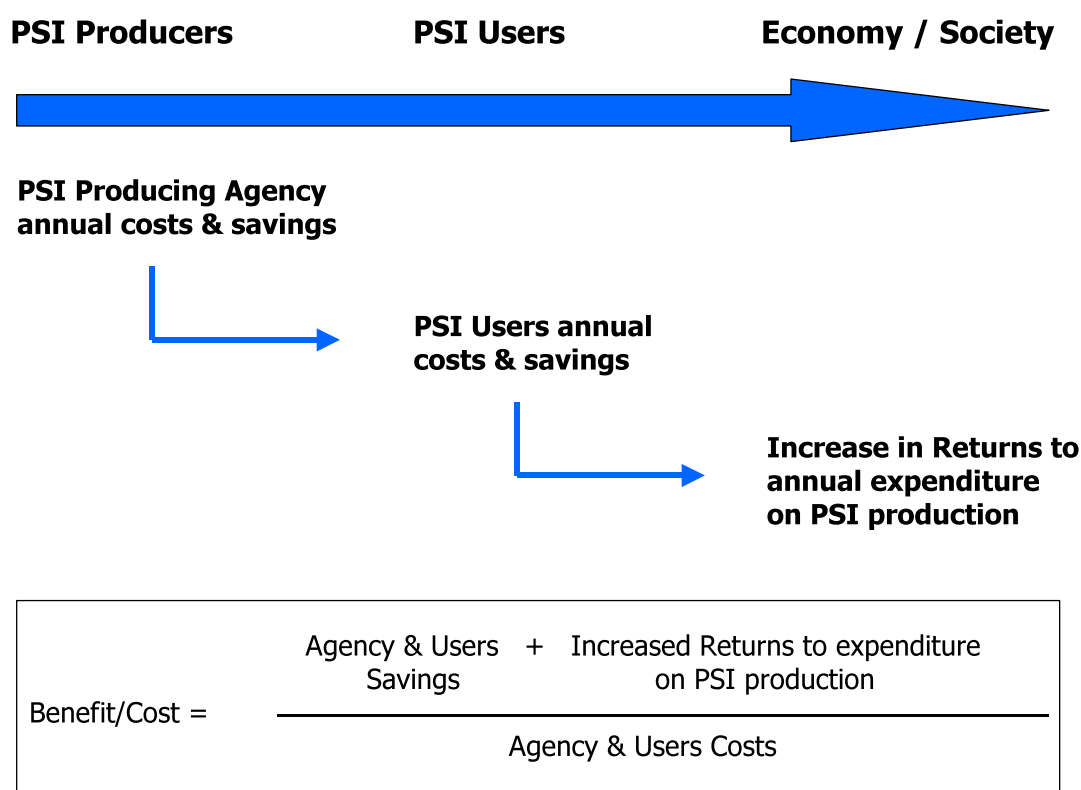
## 2 A framework for estimating cost-benefit

There are many ways in which the provision of more open access to PSI can impact upon the costs of government agency producers and the many existing and potential users of the information. This study focuses on three main elements:

- The costs and cost savings experienced by the PSI producing agencies involved in provision of free and open access to information;
- The costs and cost savings experienced by the users of PSI that relate to accessing, using and re-using the information; and
- The potential wider economic and social impacts of freely accessible PSI, measured in terms of returns to investment in its production.

Hence, the cost-benefit model outlined in this study compares the costs and benefits of providing and gaining open access to PSI against alternative models for provision (*e.g.* cost recovery). It includes: PSI producing agency costs and cost savings; PSI user costs and cost savings; and wider economic impacts in the form of increased returns to expenditure on PSI production.

Figure 1 A framework for estimating cost-benefit



Source: Author's analysis.



While there are some, typically relatively minor, one-off costs for agencies involved in the change to open access, most are recurring annual costs (*e.g.* agency IT and hosting costs, revenues foregone, etc.). Hence, the agency and user costs that are modelled are annual costs, and the cost savings annual savings. In terms of the wider benefits of open access to PSI, returns to investment in data production are recurring annual returns, lagged and discounted over the useful life of the data using a perpetual inventory method. Consequently, the cost-benefit comparisons outlined herein include annual agency and user costs and cost savings as well as the wider benefits arising from increased returns to annual expenditure on data production. They compare the costs and benefits at the time of the transition to open access (*i.e.* at the prices and levels of activity at the time).

Looking first at PSI producers and users, we suggest that it can be useful to think about possible cost impacts in terms of agency and user activities (*i.e.* an activity cost model). Table 1 shows some of the main areas in which the activity costs of PSI producers and users may be effected by providing data free (*i.e.* at zero cost or at the marginal cost of transfer), using standardised licensing that is as unrestrictive as possible (*e.g.* Creative Commons), and using standard (open) data formats.

## 2.1 Agency costs and cost savings

Agency costs and cost savings will vary from case-to-case, but there are many common elements.

**Collection / Creation:** In general, making data freely available will have little or no impact on the activities and costs associated with data collection / creation. However, there can be some impacts on what is collected if making information freely available reduces the feedback from users as to what is of most value to them. Agencies may also collect / create less if they lose revenue from data sales and need to manage costs.

Standardised licensing and data formats are unlikely to have any material impacts on agency collection costs once the standard systems are in place.

**Data Assurance (*e.g.* quality, privacy, etc.):** Making information freely available may impact data quality positively or negatively: positively if greater exposure and use raises awareness of the importance of quality; negatively if the loss of revenue from data sales puts downward pressure on costs. Free access may increase the costs associated with managing privacy and confidentiality as access and use increases.

Standardised licensing is unlikely to have any substantial impact on data assurance, although it may be somewhat easier to make data fit for purpose when it is more clearly understood ahead of time what purposes will and will not be 'allowed'. Standardised data formats are likely to make data quality easier to handle.

**Curation (agency):** Data curation costs are unlikely to be affected by making information freely available or standardised licensing, although associating licensing conditions and licenses to particular datasets would be somewhat easier. Standardisation of data formats would be likely to make digital curation easier and reduce curation costs.



**Table 1** Activity cost impacts for producers and users

<b>Activity</b>	<b>Free</b>	<b>CC standard licensing</b>	<b>Data standard</b>
Collection / Creation	May impact what and how much is collected if less revenue from sales.	..	..
Data assurance (quality, privacy, etc.)	May impact quality if no revenue.  May be more difficult to manage privacy if made freely available.	..	May make it easier to handle.
Curation (agency)	..	..	May make it easier to curate.
Dissemination (agency)	Will reduce agency transaction costs for sales.  Hosting costs may be greater if use increases.  User support costs may increase if use increases.	Reduces license transaction costs.  User support may decline.	May reduce hosting costs.  User support costs may decline.
Permission (users)	..	Reduces license transaction costs.	..
Access (users)	Reduces cost of access.	..	..
Use	Encourages/increases use.	..	May reduce cost of use.
Re-use	..	Enables re-use.	May reduce cost of re-use.

Source: Author's analysis.

**Dissemination (agency):** Making information freely available is likely to have substantial impacts on agency dissemination costs. Agency transaction costs associated with data sales will be reduced/eliminated (*e.g.* operation of sales outlets or e-commerce functions, banking charges associated with sales transactions, operation of access controls, etc.). Data hosting costs may decrease with reduced access control and purchase association and tracking, and/or they may increase if use increases. User support costs may also decrease with fewer access issues, and/or increase if use increases.

Standardised licensing is likely to reduce agency costs associated with developing and managing licensing, and reduce license related user enquiry and support costs. Standardised data formats are likely to reduce hosting and user support costs.

## 2.2 User costs and cost savings

There will be differences in user cost impacts from case-to-case and the dividing line between agency and user cost issues will vary. Nevertheless, there are likely to be common elements.

**Permission (users):** Standardised licensing is likely to reduce the user costs related to licensing (*e.g.* easier to understand licenses, standardisation across datasets and agencies, fewer license related enquiries, reduced transaction cost in accessing / obtaining licenses, etc.). Free access and standardised data formats are unlikely to have any material impact of user permission related costs.

**Access (users):** Free access obviously reduces the users' cost of access in terms of the prices paid and the transaction costs involved in each transaction/use. Standardised licensing and data formats may also reduce user access and transaction-related costs.

**Use:** Free access is likely to encourage greater use of the information, while standardised licensing and data formats are also likely to reduce the cost of, and thereby encourage greater use.

**Re-use:** Similarly, as another form of use, free access is likely to encourage greater re-use of the information, while standardised licensing and data formats are also likely to reduce the cost of, and thereby encourage greater re-use.

## 2.3 Efficiency and productivity impacts

In addition to the more direct agency and user costs and cost savings, there is a number of possible efficiency and productivity impacts arising from free access and standardised licensing and data formats.

**For agencies,** potential efficiency and productivity impacts include:

- An increase in the level of use and uses per funding dollar;
- Enhanced performance against key performance indicators;
- Enhanced agency profile from greater use and exposure, which can result in greater appreciation and central funding, and/or bring greater demand for enhanced products and services, thereby increasing revenue; and
- Greater focus on core business activities (*e.g.* reduced shopfront and e-commerce operations).

**For users,** potential efficiency and productivity impacts include:

- The purchase price savings and savings in handling and transaction costs noted above, enabling cost reduction and efficiency gains;
- Greater licensing certainty / reduced uncertainty and freedom to use leading to more predictable investment decisions, as well as savings in licensing enquiry efforts; and
- Use of better / fuller / more detailed data, rather than settling for a lesser / cheaper substitute.

There are also some important dynamics affecting the costs and benefits of more open access to PSI. For example:

- There may be first mover costs for agencies and users, making more open access, standardised licensing and data formats less cost-effective for the pioneers than it is for the followers who can learn from the pioneers. There may also be first mover advantages, such as the kudos of being a leader. This may mean that the costs and benefits experienced by first movers will tend to understate the advantages.
- There may be efficiency curve effects, with initiatives expensive to start with but efficiency improving over time. This may mean that the costs and benefits experienced in the early years of implementation tend to understate the longer-term advantages.
- There may be timing impacts and costs, as it is important how quickly data are available and how quickly licensing can be understood and affected (*e.g.* immediacy of access is important in use and avoids potentially costly delays and workflow/decision making interruptions).
- There may also be timing impacts relating to learning, as understanding of the conditions of access and licensing become more familiar to users over time and as more agencies embrace more open access and standardised licensing and data formats. Again, this may mean that the costs and benefits experienced in the early years of implementation tend to understate the longer-term advantages.
- There may also be scale economies, with greater benefits accruing as an agency's and/or user's suite of activities fall under the more open access and standardised regime. These can be significant, and may mean that the costs and benefits experienced in the early years of implementation tend to understate the longer-term advantages.

## 2.4 Wider economic impacts and benefits

Wider economic and social impacts relate to what Beagrie *et al.* (2010) referred to as investors and society or the public at large.<sup>24</sup> Clearly the funders of the data collection (investors) have a strong interest in protecting and maximising the return on their investment in the data collection / creation activity through ready availability and increased use. Where the activity is taxpayer-funded, that interest is society wide.

In the longer term, there may also be unforeseen uses and re-uses that simply cannot be accounted for, and again this may mean that the costs and benefits experienced in the early years of implementation tend to understate the longer-term advantages. Use and re-use can also have wider impacts, in terms of innovation and the development and introduction of new products, services and processes that, in turn, generate new economic activity, new business opportunities, better informed and potentially better government and business decisions.

## Box 2 Environmental impact assessments in Europe

Implementation of the European directives on Environmental Impact Assessment (EIA) and Strategic Environmental Assessment (SEA) provides an example of the potential benefits to be gained through full and open data sharing. A recent report published by the Joint Research Centre of the European Commission reported that:

- The key findings of the 2002 survey revealed that the most frequent problem that practitioners faced when using spatial data related to access (70%), followed by difficulties with finding out which data is available (56%) and unavailability of the data needed (51%). These problems were felt to increase both the time needed to prepare environmental reports by 7% to 8% and their cost by 5% to 6%. Vanderhaegen and Muro (2005) estimated that between 20,000 and 38,000 EIAs and 6,000 to 10,000 SEAs were carried out every year in the EU-25 countries. Based on this assumption and on the above mentioned results, in terms of turnover and time per study, it was estimated that annual savings of EUR 100 million to EUR 230 million per annum would be achieved if additional costs and time due to problems with the use of spatial data were removed.
- The 2009 survey found that many EIA (37%) and SEA (40%) practitioners, declared that data acquisition costs represent between 5% and 10% of the total cost to produce an EIA or SEA report. The most frequent problems practitioners face with the use of spatial data relate to finding the data (59%) and low quality (58%). These problems were followed by accessing the data (53%), integrating it (53%) and its high cost (48%). The most common concern faced by respondents was that more time was needed to prepare EIA/SEA reports (68%), followed by higher uncertainty of impacts (50%), lower accuracy and higher cost (both at 47%). The results were that practitioners needed more time to prepare EIA/SEA reports and were less certain of the impacts. The increase in costs and time caused by problems connected with the use of spatial data was between 10% and 20% which, based upon an average study cost of EUR 40,000, could lead to conservative savings in the order of EUR 150 million per annum in the countries of the European Union.

Source: Craglia, M. *et al.* (2010) *The Use of Spatial Data for the Preparation of Environmental Reports in Europe*, JRC-IES, EUR24327 EN – 2010; and GEO (2010) *GEOSS Data Sharing Action Plan*. GEO-VII Plenary.

## 2.5 A cost-benefit model

Cost-benefit analysis of PSI case studies can be founded on an activity cost model, adjusted for the specific case, which relates to the situation before the data were made freely available and afterwards, highlighting activity and cost differences for both the PSI producing agency and its users.

For the PSI producing agency, it is necessary to explore: the costs involved in data provision (*i.e.* looking at the current costs and the costs of the previous system, exploring the differences, where the costs fall and how they are met); and the agency's perceived and/or measured benefits from open provision (*i.e.* looking at the direct activity costs and/or cost savings, and cost and other impacts on the agency, as well as cost differences and where savings arise and who benefits).

For users of the PSI producing agency's data, it is necessary to explore: the access and use costs involved currently and in the previous system (*i.e.* indicative user costs of access and use in the

current and previous systems); and extrapolate from these, based on previous and current user counts.

While there are a number of approaches to assessing the wider benefits of enhanced access to PSI, here we focus on two. The first uses a simple microeconomic welfare approach and might be interpreted as indicative of the lower bound impact, and the second uses a macroeconomic approach in an attempt to include the multiplier impacts that are not captured in the first. While no more than exploratory, both have the merit of simplicity and being relatively undemanding in terms of data requirements. Briefly, they can be described as follows.

**The Welfare Approach:** Consumer surplus is the difference between the price that the consumer is willing to pay and the price actually paid (*i.e.* the net economic benefit to consumers). The level of consumer surplus depends on the price elasticity of demand, which is the percentage change in demand resulting from a percentage change in price. While there are many limitations when the price change is large and the price approaches zero, consumer surplus can be estimated from data on revenue and the elasticity of demand (*i.e.* the demand response to changes in price).<sup>25</sup>

**The Returns Approach:** There are similarities between some forms of observation and survey-based PSI and the publications and data arising from publicly funded research. To that extent, in some cases, it may be possible to use methods relating to returns to R&D to explore the wider impacts of making some forms of PSI freely available online (Houghton 2009).<sup>26</sup> The standard Solow-Swan model makes some key simplifying assumptions, including the assumption that all knowledge is equally accessible to all entities that could make productive use of it. Obviously, this is not realistic. In the real world, there are limits and barriers to access. Houghton and Sheehan (2009) developed and applied a modified Solow-Swan model in which they introduced *accessibility* as a negative or friction variable, and then looked at the impact on returns to R&D of reducing the friction by increasing *accessibility*.<sup>27</sup> While there are limitations when applying average returns to specific forms of data, indicative changes in return on investment can be estimated from expenditure data and the elasticity of demand.

These approaches are further described in Box 3 and used in the case study analyses (below).

## 2.6 Guide to data requirements

Background information required for analysis includes an understanding the PSI producing agency. For example, it may be important to:

- Clarify exactly what is and what is not freely available (free *gratis* and free *libre*), and to understand the mix and shares of each, costs of and revenues from each;
- Consider the issue of incentives for the agency of the various charging policies and agency funding (*e.g.* is revenue foregone lost or will government increase funding to compensate?);
- Explore agency revenue models (*e.g.* ‘freemium’ – free basic and charged premium), and what the relationship is between free and priced products and services (*e.g.* is ‘freemium’ revenue maximising?); and

- Examine agency revenue trends from all possible sources.

For costs, it is necessary to get a full costing of all the relevant PSI-related activities of the producing agency and of (typical) users, seeing costs as including the money and time spent on the activities, materials and overheads. For example, this might include:

- Costs of data and access to the data for users (use) and for the providing agency (provision);
- Costs of data and data access related activities for users (use) and for the providing agency (provision);
- Costs of a paid and/or licensing system, including: time spent on licensing requests (users and providers), and the number of requests processed, time spent and transaction costs of priced access (users and providers), and the number of transactions;
- Revenues gained and/or foregone, for the data directly and for other agency and/or related products and services;
- Impacts on business (*e.g.* freemium issues, such as to what extent making the PSI freely available affects priced business lines); and
- Agency incentives (*e.g.* level of collection if supporting revenue is lost, possible impacts on quality, etc.).

For use, it is necessary to understand how the change to making the PSI freely available affected demand and use. The key data required will be usage trends, in terms of visits, inquiries, purchases, datasets accessed, downloads, hits, etc. It will be necessary to:

- Tease out the issues of format (print versus digital online) from those of revenue model (priced versus free), so it is necessary to know use trends of both online and print materials and priced and free materials (*i.e.* longer-term usage trends);
- Consider adjusting these usage data for other factors and wider trends, such as:
  - The impacts of online volumes on use to explore intensification versus extension of use (*e.g.* adjust for downloads per item available);
  - The context of increasing downloads of everything (*e.g.* national or other trends in the volume of data downloaded by Internet users); and
  - The influence of information cycles on usage trends, such as special PSI data release cycles (*e.g.* ABS Census release cycles), government cycles (*e.g.* pre/post elections), general business cycles, policy topic cycles, etc.; and
- Understand the use and uses of the data, such as:
  - Whether its value realised in the year of its production or over a longer time, and if over time, how long and how is the use/value distributed;
  - What share of PSI sales/downloads go to government versus the share to private sector users/re-users;

- What share of the PSI goes to end users versus the share going to intermediaries for re-use;
- Market revenues of users, and the contribution of PSI as an input (*e.g.* share of PSI in input costs/activities); and
- The value-added of users, and the contribution of PSI as an input (as above).

For basic measures of the value of the PSI, it is necessary to explore:

- The cost of production (collection/generation) of the PSI (*i.e.* minimum value / investment value);
- Estimates of price elasticities of demand by PSI type and domain (*e.g.* to use for estimates of consumer surplus); and
- Estimates of users' willingness to pay by PSI type and domain (for non-rivalrous goods it is the sum of individual valuations).

Of course, this is a “wish list”. It is unlikely that any agency will have all of the data necessary for a complete analysis. The elements are outlined as a guide.

### 3 Public sector information case studies

Public Sector Information (PSI) is any kind of information that is produced and/or collected by a public body and is part of the institution's mandated role. It is common to differentiate between public sector information and public content. The first category comprises the public sector's knowledge, which may be the basis for information-intensive industries that use the raw data to produce sophisticated products. The second refers to cultural, educational and scientific public knowledge, for which wide public diffusion and long-term preservation (e.g. via museums, libraries, schools) are major governmental objectives (OECD 2006).<sup>28</sup> It is generally accepted that the most economically significant forms of PSI are geospatial, meteorological, hydrological and environmental information, and economic and social statistics, as they provide the greatest opportunities for commercial use and re-use and provide fundamental information in support of private and public sector decision making (Figure 2).

In 2002, the Productivity Commission released a report from an inquiry into *Cost Recovery by Government Agencies*.<sup>29</sup> The Commission concluded that cost recovery should be implemented for economic efficiency reasons, not merely to raise revenue. Information agencies should define a basic product set according to public good characteristics, significant positive spillovers and other Government policy reasons. The basic product set should be funded from general taxation revenue. Additional information products should be classified into three broad categories and priced accordingly: (i) dissemination of existing products at marginal cost; (ii) incremental products (which may involve additional data collection or compilation) at incremental (avoidable) cost; and (iii) commercial (contestable) products according to competitive neutrality principles. Cost recovery should not be implemented where it is not cost effective, it would be inconsistent with policy objectives, or it would unduly stifle competition and industry innovation (for example, through 'free rider' effects).

Following the Productivity Commission Inquiry, the Australian Government adopted a formal cost recovery policy in December 2002, in order to improve the consistency, transparency and accountability of Commonwealth cost recovery arrangements and promote the efficient allocation of resources.<sup>30</sup> The guidelines required fees and charges set by Government agencies to reflect the costs of producing and providing the products and services. In addition, where Government agencies produce products or services in direct competition with private sector providers, the guidelines required that the prices be set to reflect commercial costs and to ensure that the Australian Government's Competitive Neutrality principle are observed.

The cost recovery policy was updated in 2005, and government agencies were required to undertake a review and prepare cost recovery impact assessments. The policy stated that:

- Products and services funded through the budget process form an agency's 'basic information product set' and should not be cost recovered; but
- Commercial, additional and incremental products and services that are not funded through the budget process fall outside of an agency's 'basic product set' and may be appropriate to cost recover.<sup>31</sup>



Figure 2 PSI and public content domains with examples

<div>Commercial Re-Use PSI</div> <div>↑</div> <div>↓</div> <div>Making Available PSC</div>	Geographic Information	Cartographic Land Use (Cadastral) Spatial / Geographic Admin and Political Boundaries Topographical Elevation
	Meteorological and Environmental Information	Oceanographic Hydrographic Environmental (quality) Atmospheric Meteorological (weather)
	Economic and Business Information	Financial Company Economic Statistics Industry and Trade
	Social Information	Demographic Attitude Survey Health Education and Labour Force
	Traffic and Transport Information	Transport Network Traffic Transport Statistics Vehicle Registration
	Tourist and Leisure Information	Hotel Tourism Entertainment
	Agriculture, Forestry and Fisheries Information	Cropping and Land Use Incomes and Resources Harvest Livestock
	Natural Resources Information	Biological and Ecological Energy and Resource Consumption Geological and Geophysical
	Legal System Information	Crime and Convictions Laws Rights and Duties Legislation Judicial Decisions Patents and Trademarks
	Scientific Information and Research Data	University Research Publicly-Funded Research Governmental Research
	Educational Content	Academic Papers and Studies Course and Lecture Material
	Political Content	Government Press Releases Government Proceedings Green Papers
	Cultural Content	Museum and Gallery Archeological Sites Library Resources Public Service Broadcasts Archives

Source: OECD (2006) *Digital Broadband Content: Public Sector Information and Content*, OECD: Paris.

General principles included that material in the basic information set be made freely available online, packaged material be priced at no more than the marginal cost of transfer, and customised material priced at a level not exceeding the full cost of transfer. Hence, as the Government 2.0 Taskforce noted:

*Australia has been moving towards more open data management since at least 2001, which saw the Spatial Data Access and Pricing Policy, making Australia a leader in deliberately seeking economic and social benefits by moving away from selling data towards free distribution. The Australian Bureau of Statistics (ABS) followed suit in the*

*mid 2000s. Today both ABS and Geoscience Australia (GA) are again leading, both licensing much of their output using ‘Creative Commons’ attribution only and thereby permitting others to use, and remix it with minimal cost and restriction.*<sup>32</sup>

The following sections explore the cases of the Australian Bureau of Statistics (national statistics), GeoScience Australia (spatial data) and the National Water Commission & Bureau of Meteorology (hydrological data). Due to data limitations and issues with timing the latter two case studies are incomplete.

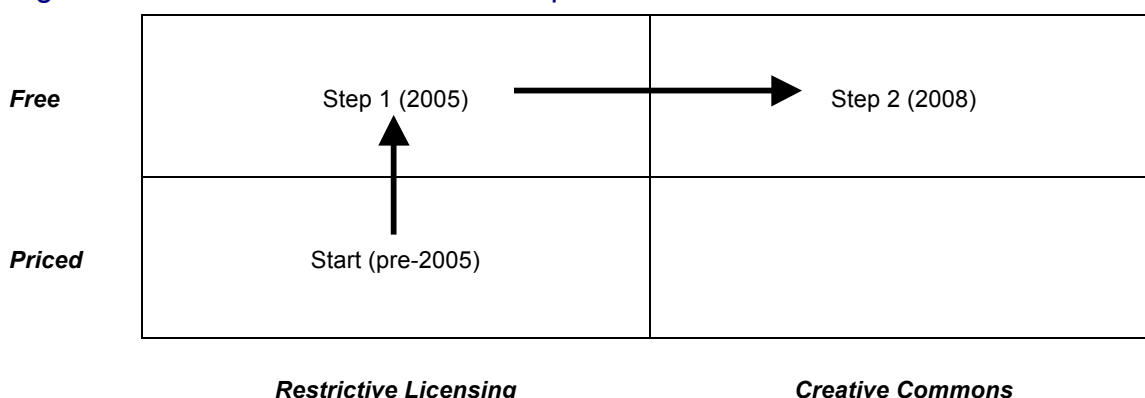
### 3.1 National Statistics (Australian Bureau of Statistics)

Following a review of its charging practice against new Government cost recovery policy guidelines, the ABS divided its statistics into:

- A Basic Information Set (BIS), which includes an extensive range of statistics for the wider Australia community and is funded by taxpayers and provided free of charge; and
- An Additional Information Set (AIS), which includes the ABS Supplementary Information Set (SIS) for which pricing is based on full cost recovery, and Commercial Information Set (CIS) for which pricing follows the Competitive Neutrality Principle.

In June 2005, the ABS obtained additional funding from the Australian Government to enable free access to ABS publications on its website. In December 2005, the Minister made the announcement that all ABS statistical output on the website (*i.e.* both publications and data) would free of charge (Tam 2009).<sup>33</sup>

**Figure 3 ABS moves towards open access**



Source: ABS consultations. Author's analysis.

As recent Web 2.0 technologies increase the potential to use, share and ‘mix and match’ data to add value to information, the ABS saw the need to have an internationally recognised licensing framework for accessing, using and re-using its data. In December 2008, the ABS introduced

Creative Commons licensing, adopting the Attribution 2.5 Australia licence for material on the ABS website (Tam 2009).<sup>34</sup>

In consultation with the ABS, their path towards Open Access was described in terms of a two-by-two matrix of pricing and licensing (Figure 3), in which the ABS moved towards open access as shown (*i.e.* moving on pricing in 2005 and licensing in 2008). It is felt that the next/current challenge for accessibility is standardisation on (open) file formats and standards.

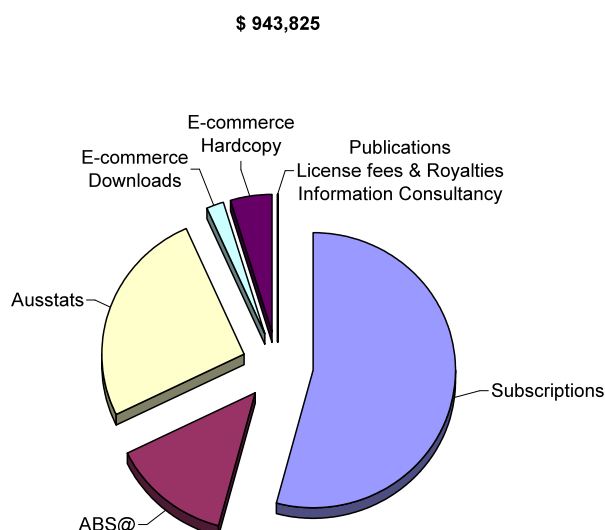
In the following sections we explore the costs and benefits involved in making ABS publications and statistics freely available online and Creative Commons licensing.

### 3.1.1 Agency costs and benefits

Looking at the direct costs and cost savings experienced by the agency, the ABS report estimated annual cost savings from the move to *making publications and statistics freely available online* of around \$945,000 and annual revenue lost of around \$4.5 million circa 2006-07.

Significant savings were realised in staff and other costs in the operation of subscriptions, ABS@, Ausstats and regional office advice/support (Figure 4). Savings from e-commerce activities were limited to transaction fees in the early years as the services continued. With decommissioning, however, additional annual savings were realised. There were no savings relating to publications (*e.g.* bookshop operations), information consultancy, license fees and royalties as services continued (although the latter is now discontinued).

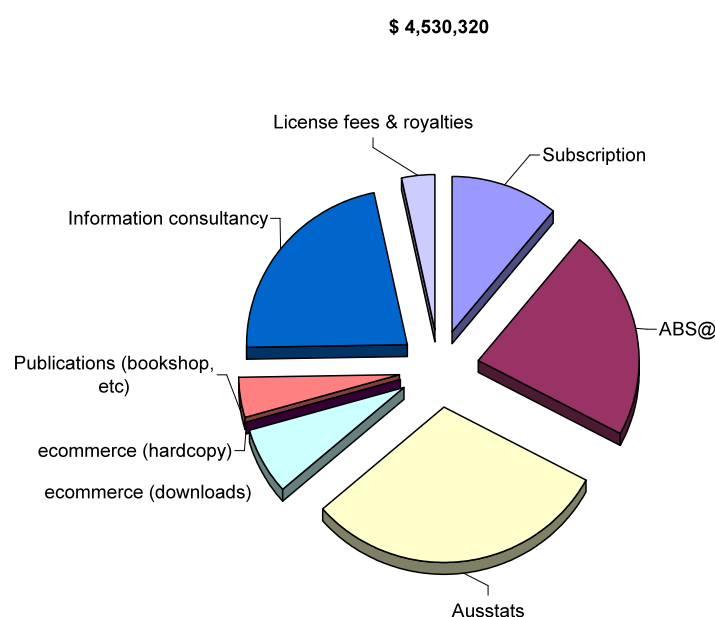
Figure 4 Estimated annual savings from FPOW/FSOW (Circa 2006-07)



Note: FPOW = Free Publications on the Web, and FSOW = Free Statistics on the Web.  
Source: ABS consultations. Author's analysis.

Comparing 2005-06 and 2006-07 revenues with those from 2004-05 suggests that revenue losses from data sales were substantial, with annual revenue foregone from Ausstats of around \$1.3 million, ABS@ of around \$1 million and subscriptions of around \$490,000. Information consultancy revenues also fell by around \$1 million a year (Figure 5). Hence the net position resulting from the free provision of publications and statistics online for the ABS was an annual loss of around \$3.6 million.

**Figure 5**      **Estimated ongoing annual loss in revenue due to FPOW/FSOW (Circa 2006-07)**



Note: FPOW = Free Publications on the Web, and FSOW = Free Statistics on the Web. Ongoing annual losses estimated by setting 2005-06 and 2006-07 revenues against the benchmark year 2004-05.  
Source: ABS consultations. Author's analysis.

In May 2005, it was announced that the Government would provide additional funding to the ABS of \$76.2 million over four years “to strengthen and expand statistical collections and to improve the availability of data to the community.”<sup>35</sup> In the 2005-06 Budget, the ABS received around \$11 million per annum for four years to support upgrades.<sup>36</sup> It has not been possible to determine what share of this money was aimed at improving the availability data (*i.e.* effectively compensating for losses sustained as a result of the free provision of publications and statistics online). Consequently, the welfare analysis (presented below) does not include any possible deadweight loss associated with taxation.

Estimated annual direct staff cost of providing licences for on-provision of ABS customised data at 2008 was around \$125,000. Revenue foregone by *the introduction of Creative Commons licensing* was around \$115,000 for customised data outputs. In addition, there were cost savings of around \$47,000 in providing licensing education, advice and support to external clients, but

some of these costs have shifted to internal activities. Savings from discontinuing use of ABS@ containers could not be ascertained. Hence the net position for the ABS from CC licensing was likely to have been close to neutral.

Examining the impact of the adoption of Creative Commons Licensing in an internal document, the ABS noted that:

*The main impact of licensing customised data under Creative Commons is to allow more data sharing and freedom of reuse. This will ensure freer and more widespread use of ABS data, and facilitate innovative use of ABS data for the public good. Licensing customised data under Creative Commons: meets public expectations with regard to open government; supports increased access to and use of statistics; facilitates data sharing (including across government); allows for more timely reuse of statistics; simplifies the licensing process; clarifies licence conditions; makes sense to a growing percentage of people who recognise and understand CC licence symbols and conditions; and facilitates innovation.*

*The one-off costs involved include: managing the project to introduce the new licensing regime (but no ongoing licence administration costs for data licensed under CC); establishing policy guidelines to determine when data is too sensitive to be licensed under CC and providing alternative licence(s); updating operating procedures and business systems to ensure that the need for exceptions to the standard CC license is considered for all information consultancy jobs; and providing training and documentation for information consultancy staff about the new licensing arrangements and protocols for identifying and handling exceptions.*

*The on-going cost of determining whether or not the more restrictive licence is applicable, and then providing it where necessary, can be absorbed into the normal running costs for information consultancy services.*

*Productivity Savings include: current and projected annual revenue from secondary distribution licences; current staff costs to administer secondary distribution and provide free licences; removal of secondary distribution licences is revenue-neutral (on an ongoing basis), but does free up staff resources for core and strategic business (resulting in savings); and saving in no longer needing to provide free licences. By no longer requiring these licences, ABS resources will be freed up to be used in a more productive way.*

*There are also indirect productivity gains with respect to the involvement of other sections in the process (e.g. Accounts Payable and Receivable, Procurement Services and Contract Support) and the input of senior staff in an advisory capacity from time to time. These savings, though real, are difficult to quantify.*

*...If we license the customised data delivered via consultancies under Creative Commons, there will be no licences to manage in-house and we therefore do not need to charge for licence administration. While this may look like a loss of revenue, it is balanced by the fact that the task no longer needs to be done.<sup>37</sup>*

**Table 2** Estimated net position for ABS from FPOW/FSOW and CC licensing (per annum)

	<i>Value</i>
<b>Estimated net position for ABS from FPOW/FSOW:</b>	
Estimated annual savings	943,823
Estimated annual loss	4,530,321
Net revenue position (loss)	-3,586,498
<b>Estimated net position for ABS from CC licensing:</b>	
Estimated annual savings	171,667
Estimated annual loss	115,000
Net revenue position (gain)	56,667
<b>Total net position</b>	<b>-3,529,832</b>
Benefit / Cost ratio	0.24

Note: FPOW = Free Publications on the Web, and FSOW = Free Statistics on the Web.

Source: ABS consultations. Author's analysis.

Overall, the estimated direct agency impacts of making publications and statistics freely available online and using Creative Commons licensing for customised data resulted in an annual loss of around \$3.5 million (Table 2).

### 3.1.2 User costs and benefits

While it is impossible to collect detailed information from the many and varied users of ABS data, it is possible to explore a number of proxy indicators of use and develop some estimates of possible users' costs and benefits.

The mirror image of agency savings in transaction costs and lost revenue is a guide to users' transaction cost savings and reduced data costs. Table 3 presents approximate estimates of direct annual user cost savings. While these are estimates, it is immediately apparent that the likely direct users' savings of around \$5 million per annum are greater than the agency's net cost (loss) position of \$3.5 million per annum.

**Table 3**      **Estimated annual user cost savings from free publications and statistics online and standardised CC licensing**

	<i>Parameter</i>	<i>Value</i>
<b>Assume:</b>		
Cost of person per hour incl. overheads and on-costs		75
<b>Impacts of free:</b>		
<i>e-commerce transactions</i>		
Time (mins)	10	13
Bank costs (same as ABS)	0.9	18,823
Transactions (2006-07)	20,914	261,425
Total annual		280,248
ABS annual revenue loss is user saving (cash)		4,530,321
<b>Estimated users' savings (per annum) of free on the web</b>		<b>4,810,569</b>
<b>Impacts of CC:</b>		
Time saved on license enquiries		41,667
License cost savings		115,000
<b>Estimated users' saving (per annum) of CC licensing</b>		<b>156,667</b>
<b>Total estimated users' savings</b>		<b>4,967,235</b>
Source: ABS consultations. Author's analysis.		

### 3.1.3 Indicators of use (web statistics)

An important source of proxy indicators of use is ABS website statistics, which show something of the impacts of making publications and statistics freely available online. However, there are a number of limitations and issues to consider when exploring website statistics and trying to isolate the impacts of making publications and data freely available from other trends and influences. These include the impacts of:

- General trends in the use of online materials (*e.g.* the mid-2000s saw an increase in downloads across the board);
- Annual and cumulative publication volumes, including the Census publication cycle (*e.g.* there are likely to be more downloads as more is made available to download);
- Changes to other related factors (*e.g.* pricing, policy cycles, etc.); and
- Changes in the way in which ABS website hits and downloads were recorded and reported.

In the following paragraphs we attempt to explore the extension and intensification of use as best we can in view of these issues and limitations.

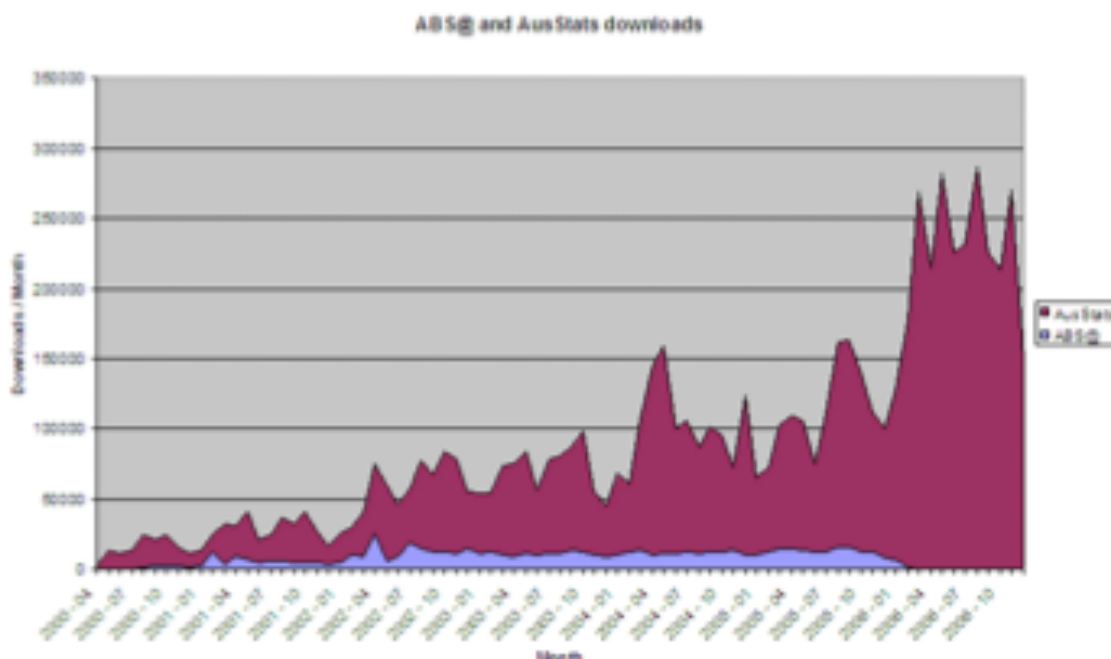
ABS (2007), Fitzgerald (2007) and pwc (2010) noted a surge in downloads from the ABS website following the introduction of free publications and statistics online.<sup>38</sup> ABS (2007) reported that:

*There was a very significant increase (+21 million hits) in ABS website traffic in 2006, which represents an increase of 34% over 2005. Historical trend data for [the] ABS@ subscriber environment indicates that the transition of subscribers to the main website has had very little impact on the overall upward trend.*

*There was a 106% increase in the number of downloads from the ABS website in 2006. However, in the context of the overall traffic increase, this represents an increase of only 0.9% (from 1.8% to 2.7%) in the ratio between hits and downloads, and the gap between hits and downloads has widened significantly.<sup>39</sup>*

The chart below (Figure 6), showing downloads per month from 2000 to 2006 has appeared in a number of reports and presentations.<sup>40</sup>

Figure 6      ABS@ and Ausstats downloads



Source: ABS (2007) *Impact of Free Stats on the Web on ABS Website Usage*, Canberra.

Up to and including 2008-09, page views and products downloaded were reported as shown in Table 4, with 2006-07 being the first full year of free downloads and 2007 being a census release year.



**Table 4** ABS website usage, 2003-04 through 2008-09

	<b>Page Views</b>	<b>Percentage change (page views)</b>	<b>Products Downloaded</b>	<b>Percentage change (downloads)</b>
2003-04	48,383,816	..	948,956	..
2004-05	60,573,254	25%	962,872	1%
2005-06	78,054,933	29%	1,868,280	94%
2006-07	101,693,436	30%	4,501,530	141%
2007-08	140,058,970	38%	7,029,854	56%
2008-09	193,519,053	38%	3,150,630	-55%

Source: ABS Annual report 2008-09. Author's analysis.

The ABS *Annual Report 2009-10* reported revised data, noting that:

*Data for pages viewed and products downloaded have been rebased. Rebased refers to the process by which ABS uses the most recent results (in this case, 2009–10) to revise all prior estimates. Rebased ensures that the time series are comparable with the latest estimates made using the latest available tools. The rebasing of page views embraces current web analytic methodologies in that it excludes the influence of internal staff web access, indexing/spidering and robots, and single-session page revisits. The rebasing of downloads was necessary due to the decommissioning of legacy infrastructure and implementation of a new web analytics tool.<sup>41</sup>*

The re-basing appeared to involve simple multiples of 3.8 for page views and 2.1 for downloads, so the re-basing has no impact on growth rates. There were three other important notes to the revised data table: pages published in 2006-07 included approximately 100,000 Census first release pages; the 2007-08 and 2008-09 ABS Annual Reports published the number of pages published in 2007-08 as 193,515, which was a cumulative total and has been corrected in this edition; and the 2008-09 ABS Annual Report published the number of pages published in 2008-09 as 238,886, which was a cumulative total and has been corrected in this edition. Taking these issues into account the revised data are as shown in Table 5.

**Table 5** ABS website usage, 2003-04 through 2009-10 (re-based)

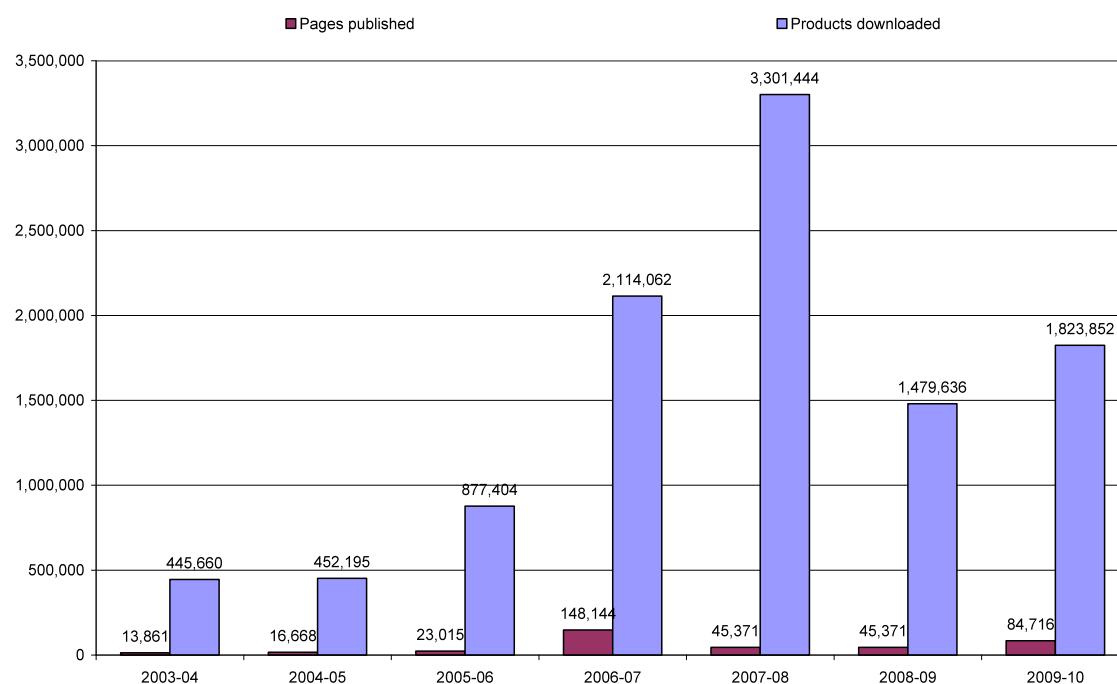
	<b>Page Views (number)</b>	<b>Pages Published (number)</b>	<b>Products Downloaded (number)</b>	<b>Page Views (% change)</b>	<b>Pages Published (% change)</b>	<b>Products Downloaded (% change)</b>
2003-04	12,727,980	13,861	445,660	..	..	..
2004-05	15,934,567	16,668	452,195	25%	20%	1%
2005-06	20,533,354	23,015	877,404	29%	38%	94%
2006-07	26,751,755	148,144	2,114,062	30%	544%	141%
2007-08	36,844,298	45,371	3,301,444	38%	-69%	56%
2008-09	50,907,654	45,371	1,479,636	38%	0%	-55%
2009-10	54,983,668	84,716	1,823,852	8%	87%	23%

Note: Re-basing for 2003-04 and 2004-05 done by the author using the same multiples.

Source: ABS *Annual Report 2009-10*. Author's analysis.

Apart from the Census release cycle and a general increase in use of online materials, it is difficult to discern any particular trends in these use data (2006-07 being the first full year of free downloads and 2007 being a Census release year). Products downloaded in 2009-10 were 18% higher than the annual average throughout the period and pages published 36% higher, although it is clear that product downloads have more than trebled since the transition to free online publications and data (Figure 7).

**Figure 7** ABS website usage, 2003-04 through 2009-10 (re-based)



Note: Re-basing for 2003-04 and 2004-05 done by the author using the same multiples.

Source: ABS *Annual Report 2009-10*. Author's analysis.

To isolate use trends, we explore intensity of use in terms of downloads per page viewed, per page published (annually and cumulatively), and per release. The results are in Table 6 (Figure 8).

Again, noting that 2006-07 was the first full year of free downloads and 2007 was a census release year, it is difficult to discern any particular pattern in these intensity of use data. Nevertheless, in the next section we use these data to inform estimates of demand elasticity and, thereby, estimates of the impact of making ABS publications and statistics freely available online on consumer welfare.

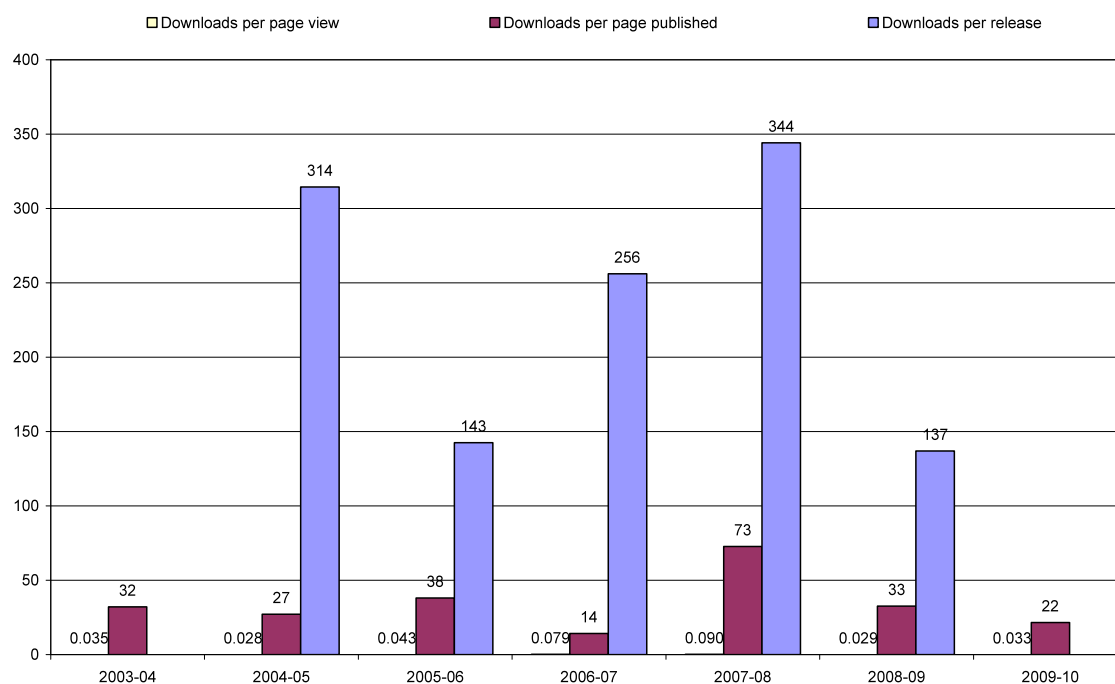
**Table 6** Downloads per page view, page published and release, 2003-04 through 2009-10

	<i>Per Page View (number)</i>	<i>Per Page Published (number)</i>	<i>Per Release (number)</i>	<i>Per Page View (% change)</i>	<i>Per Page Published (% change)</i>	<i>Per Release (% change)</i>
<b>Original data:</b>						
2003-04	0.020	68	..	..	..	..
2004-05	0.016	58	670	-19%	-16%	..
2005-06	0.024	81	304	51%	41%	-55%
2006-07	0.044	30	545	85%	-63%	80%
2007-08	0.050	155	733	13%	410%	34%
2008-09	0.016	69	292	-68%	-55%	-60%
<b>Re-based data:</b>						
2003-04	0.035	32	..	..	..	..
2004-05	0.028	27	314	-19%	-16%	-69%
2005-06	0.043	38	143	51%	41%	-55%
2006-07	0.079	14	256	85%	-63%	80%
2007-08	0.090	73	344	13%	410%	34%
2008-09	0.029	33	137	-68%	-55%	-60%
2009-10	0.033	22	..	14%	-34%	-100%

Note: Downloads per page published refers to annual publications. Downloads per estimated cumulative page published have declined steadily throughout the period.

Sources: ABS annual reports (various years). Author's analysis.

**Figure 8** Downloads per page view, page published and release, 2003-04 through 2009-10 (re-based)



Sources: ABS annual reports (various years). Author's analysis.

### 3.1.4 Wider impacts of use

In this section we explore two possible approaches to estimating the wider impacts of making ABS publications and statistics freely available online. The first uses a simple microeconomic welfare approach and might be interpreted as indicative of the lower bound impact, and the second uses a macroeconomic approach in an attempt to include the multiplier impacts ignored in the first (Box 3).

#### Box 3 Method: Welfare impacts and returns to expenditure

**Welfare Impacts:** One approach that has been used to measure the benefits of making PSI freely available is to estimate the resulting increase in consumer surplus. Consumer surplus is the difference between the price that the consumer is willing to pay and the price actually paid (*i.e.* the net economic benefit to consumers). The level of consumer surplus depends on the price elasticity of demand, which is the percentage change in demand resulting from a percentage change in price.

While there are many limitations when the price change is large and the price approaches zero, consumer surplus can be estimated from data on revenue and the elasticity of demand (*i.e.* the demand response to changes in price).<sup>42</sup> Assuming that demand is linear, following the one-half rule, the increase in consumer surplus equals revenue plus one-half of the revenue multiplied by the elasticity:

$$\Delta CS = (p_0 q_0) + \frac{1}{2} e(p_0 q_0)$$

where  $e = \Delta q/q_0$

The assumption of linearity is necessary in the face of data limitations, but is likely to lead to an underestimate of consumer surplus as most observed empirical demand relationships are convex (*i.e.* demand increases/decreases faster than linearly as price decreases/increases).<sup>43</sup>

**Returns to Expenditure:** There are similarities between some forms of PSI, including ABS publications and statistics and geospatial science and fundamental spatial data, and the publications and data arising from publicly funded research. To that extent, it is possible to use methods relating to returns to R&D to explore the wider impacts of making such PSI freely available online (Houghton 2009).<sup>44</sup>

In a number of papers and reports, Houghton *et al.* (2009; 2010) have developed and applied a modified Solow-Swan model to estimate the potential impacts of more open access to research publications and data on social returns to R&D.<sup>45</sup> The standard Solow-Swan model makes some key simplifying assumptions, including the assumption that all knowledge is equally accessible to all entities that could make productive use of it. Obviously, this is not realistic. In the real world, there are limits and barriers to access. So, we introduce *accessibility* into the standard model as a negative or friction variable, and then look at the impact on returns to R&D of reducing the friction by increasing *accessibility* (for details of the method see Houghton and Sheehan 2009).<sup>46</sup>

Returns to R&D vary widely, but a characteristic finding is that returns are high – often in the region of 20% to 60% (Bernstein and Nadiri 1991; Griliches 1995; Industry Commission 1995; Salter and Martin 2001; Scott *et al.* 2002; Dowrick 2003; Shanks and Zheng 2006; Martin and Tang 2007; Sveikauskas 2007; Hall *et al.* 2009).<sup>47</sup> To the extent that government-funded PSI producing agencies' activities can be seen as similar to research (*e.g.* producing survey-based or observational data), they may be supposed to exhibit similar returns.

Source: Author's analysis.

## Consumer surplus

Estimates of elasticity depend on ‘before’ and ‘after’ measures of demand. For the reasons noted above (*e.g.* general Internet use trends, the impacts of the Census release cycle on ABS website downloads, etc.), we must establish the before and after estimates with care. We note that:

- There are limited website data prior to ABS publications (July 2005) and statistics (December 2005) being made freely available;
- While 2006-07 is the first full year of FPOW/FSOW, it is influenced by the Census release cycle; and
- There are many other influences on download behaviour (*e.g.* general Internet use trends, shifts from print to online distribution, pricing changes, etc.), which tend to increase in their influence over time.

Pollock (2009) estimated elasticity by comparing average ABS website downloads for 2003-04 and 2004-05 with average downloads for 2005-06 and 2006-07, suggesting a demand elasticity of 2.33 (we express elasticities as absolute values throughout this report). In addition to this, we explore: (*i*) the three-year averages of 2003-04 to 2005-06 and 2006-07 to 2008-9, suggesting an elasticity of 2.88; and (*ii*) the single years 2004-05 and 2005-06, suggesting an elasticity of 0.94. We also explore intensity of use, most particularly downloads per page published, and find that comparing: (*i*) the three-year averages 2003-04 to 2005-06 and 2006-07 to 2008-9 suggests an elasticity of 0.23; and (*ii*) the single year 2004-05 and 2006-06 suggests an elasticity of 0.41.

**Table 7** Estimates of consumer surplus from ABS making publications and statistics free online

	<i>Elasticity / Value</i>
<b>Elasticity estimates:</b>	
Elasticity (from downloads per page published 3yr 2003-04 to 05-06 and 2006-07 to 08-09)	0.23
Elasticity (from downloads per page published 2004-05 to 2005-06)	0.41
Elasticity (from downloads 2004-05 to 2005-06)	0.94
<b>Consumer surplus estimates:</b>	
Consumer surplus per page published (3yr 2003-04 to 2005-06 and 2006-07 to 2008-09)	3,933,406
Consumer surplus per page published (2004-05 to 2005-06)	4,245,608
Consumer surplus raw downloads (2004-05 to 2005-06)	5,190,137

Note: Elasticities expressed as absolute values.

Source: Author's analysis.

Taking the revenue parameter to be equivalent to ABS estimates of revenue lost by making publications and statistics freely available online (less the impacts on consultancy revenues) (*i.e.* \$3.53 million per annum), the single year estimates suggest an increase in consumer surplus of some \$4.2 million to \$5.2 million per annum at 2004-2006 prices and levels of activity. Estimates using the three-year averages suggest an increase in consumer surplus of some \$4 million per annum, which while affected by the Census release cycle (hence the use of

downloads per page published), has the advantage of taking account of the time taken for users' awareness to develop and demand to adjust to the new price (*i.e.* free).

These estimates do not include any deadweight loss associated with compensating ABS for the loss in revenue through additional government funding, as it has not been possible to determine what, if any, compensation was paid. However, government is a major user of ABS publications and statistics, suggesting that the deadweight loss might be small. Nor do these estimates include producer surplus, as the data are collected for public policy purposes anyway, there would be no profit in the alternative cost-recovery model, and the revenue foregone is included in the agency and users' costs and cost savings, discussed above.

### Returns to expenditure

There are similarities between ABS publications and statistics and the publications and data arising from publicly funded research. To that extent, it is possible to use methods relating to returns to R&D to explore the wider impacts of making ABS publications and statistics freely available online.

As a first approximation, we assume a lower bound 20% return on expenditure and that 50% of the ABS's government-funded activities are devoted to generating freely available publications and statistics. Returns to research expenditure are recurring over the useful life of the knowledge, but lagged to account for the time between the conduct of research and its use impacts, and discounted accordingly. Our consultations with ABS suggested that it is reasonable to assume that ABS statistics have a useful life equivalent to the five-year Census cycle, with impacts exhibiting a normal distribution. Hence, we estimate increases in returns to a single year's expenditure over five years, discounted.

**Table 8** Estimates of increase in returns to annual spending from ABS making publications and statistics free online

	<i>Value</i>
ABS annual revenue from Government circa 2005-2006	295,000,000
Assume 50% of Government funded activity time is spent producing the data	147,500,000
A 20% return (assuming the lower bound average for research)	29,500,000
<b>Elasticities (as above):</b>	
Increase in access (from downloads per page published 3yr 2003-04 to 2005-06 and 2006-07 to 2008-09)	23%
Increase in access (from downloads per page published 2004-05 to 2005-06)	41%
Increase in access (from downloads 2004-05 to 2005-06)	94%
<b>Increases in returns:</b>	
Increase in returns (from downloads per page published 3yr 2003-04 to 2005-06 and 2006-07 to 2008-09)	10,472,226
Increase in returns (from downloads per page published 2004-05 to 2005-06)	18,583,295
Increase in returns (from downloads 2004-05 to 2005-06)	43,122,316
Source: Author's analysis.	

The impacts of the observed increases in use, as indicated by ABS website downloads, on average social returns to expenditure are estimated as described (Box 3), with the single year estimates suggesting an increase in social returns of some \$18 million to \$43 million at 2004-2006 prices and levels of activity (Table 8).

Following an extensive review of the literature and evidence, Pollock *et al.* (2008) suggested that their welfare analysis might be ignoring PSI-related multiplier effects that they estimated to be between 1.5 and 7.0, with a mid-point of 3.0. Our simplified analysis suggests a consumer surplus of around \$4.2 million per annum and increased social returns of around \$18 million (4.4 times).

### 3.1.5 Summary of impacts

Noting that these estimates are no more than preliminary and indicative, putting the elements

**Table 9** Summary of ABS case study annual costs and benefits (circa 2005-06)

	Value
<b>Agency impacts:</b>	
Net impact of FPOW/FSOW	-3,586,500
Net impact of CC licensing	56,700
Total net position for ABS	-3,529,800
<b>User impacts (direct):</b>	
Net impact of free	4,810,600
Net impact of CC licensing	156,700
Total users' savings	4,967,200
<b>Welfare estimates:</b>	
Surplus per page published (2003-04 to 2005-06 and 2006-07 to 2008-09)	3,933,400
Surplus per page published (2004-05 to 2005-06)	4,245,600
Surplus raw downloads (2004-05 to 2005-06)	5,190,100
<b>Returns estimates:</b>	
Increase in returns (downloads per page published 2003-04 to 2005-06 and 2006-07 to 2008-09)	10,472,200
Increase in returns (downloads per page published 2004-05 to 2005-06)	18,583,300
Increase in returns (downloads 2004-05 to 2005-06)	43,122,300
<b>Overall impacts estimates:</b>	
Estimated total costs	4,645,300
Estimated total benefits (likely lower bound)	6,082,700
Estimated total benefits (incl. multiplier/spillovers)	24,666,000
Benefit/cost ratio (likely lower bound)	1.3
Benefit/cost ratio (incl. multiplier/spillovers)	5.3
Source: Author's analysis.	

together suggests that the cost of making ABS publications and statistics freely available online and CC licensing amounted to around \$4.6 million per annum circa 2005 and that the benefits of doing so may have amount to around \$6 million to \$25 million. Hence the benefits may be around 1.3 to 5.3 times the costs (Table 9).

## 3.2 Spatial Data (Office of Spatial Data Management & Geoscience Australia)

In September 2001, the Australian Government launched the Commonwealth Policy on Spatial Data Access and Pricing (SDAP) to provide access to spatial data free of charge online, or to provide packaged data at a charge of no more than the cost of transfer. It was introduced in February 2002, and applies to all Australian Government departments and agencies that provide spatial data. The purpose of the pricing policy is to provide a whole-of-government approach to pricing of fundamental spatial data, and it applies to data collected by agencies in the public interest. The policy is premised on the view that all fundamental spatial data should be freely available at marginal cost of transfer in order to maximise the net economic and social benefits arising from its use.<sup>48</sup>

Under the pricing policy, data should be made available in one of three forms:

- Online spatial data should be made available free, as soon as appropriate technology becomes available within the custodian agency;
- Packaged spatial data should be made available at a price not exceeding the marginal cost of transfer; and
- Customised spatial data should be made available at a price not exceeding the full cost of transfer.

The policy also adds that there will be no restrictions on commercial use or value-added activities related to the spatial data, although copyright may be reserved by the Commonwealth and each transaction will be covered by a licence setting out the conditions of the transfer.<sup>49</sup>

In July 2008, the Spatial Data Management Group agreed to adopt the Creative Commons Licensing framework to enhance public access to spatial data held by Australian Government agencies and make it available without undue restriction on commercial use or re-use for value-added activities. In announcing its decision to apply CC licences to key mapping and other information products, Geoscience Australia (GA) emphasised that the use of the “easy to understand, royalty-free, modular, off the shelf [CC] licences” would make it easier for visitors to GA’s website to use and access information. In November 2009, GA began licensing all the material on its website, and the OzCoasts website which it hosts, under the Creative Commons Attribution 2.5 Australia licence (CC-BY). In June 2010, following the launch of version 3.0 of the Australian Creative Commons licences, GA’s website was updated to state that, unless otherwise noted, all GA material on the site is licensed under the Creative Commons Attribution 3.0 Australia licence (CC-BY) (Fitzgerald 2010).<sup>50</sup>



### 3.2.1 Agency and user costs and benefits

There has been a number of studies exploring the value of fundamental spatial data and the cost-benefit and wider impacts of various pricing models. As a report using a similar microeconomic welfare approach to making an *Economic Assessment of Spatial Data Pricing and Access* was published by ANZLIC as recently as November 2010 (pwc 2010), we do not intend to repeat the analysis. Rather we report the relevant parts of the pwc analysis and add elements of our own in an attempt to complete the picture. Moreover, following a major review, there is a reorganization of spatial data access management underway at the time of writing, making original data collection from the agencies concerned difficult and somewhat reducing the relevance of reporting detailed agency costings from the former organizational structure.

#### The pwc report

In the report *Economic Assessment of Spatial Data Pricing and Access*, PriceWaterhouseCoopers (pwc 2010) used a microeconomic welfare approach to examining the relative costs and benefits of alternative models for pricing and accessing geospatial data. Rather than exploring the costs and benefits *per se*, they performed a pair-wise comparison of alternative pricing models – most relevantly, comparing the costs and benefits of providing free fundamental data relative to using cost recovery. Unfortunately, their analysis did not include possible agency or user transaction-related cost savings.

Exploring the case of Geoscience Australia (GA) topographical data, pwc noted a total cost of production circa 2009-10 of \$13.3 million (*i.e.* the revenue that would be earned in a cost recovery model), and marginal cost of distribution, based on the cost of 2008-09 web downloads, of \$1,842 or just 0.01%, and so they set marginal cost to zero. Government's share of purchases was estimated at 50%. Despite a long discussion on elasticity of demand and possible multiplier effects, pwc assume both the elasticity and multiplier to be 1.

Their computed welfare impacts of free provision of GA topographical data relative to cost recovery included:

- An agency loss of \$13.3 million per annum in revenue foregone;
- A gain for government users of \$10 million per annum, of which \$6.65 comes in access costs saved and the remainder from additional use and distributional changes in deadweight loss;
- A gain for private consumers of \$8 million per annum, of which \$6.65 comes from access costs saved and the remainder from additional use and distributional changes in deadweight loss; and
- An overall increase in net welfare of \$4.7 million per annum, as the net welfare impact on government is negative (*i.e.* \$ -3.3 million per annum).

If we extrapolate this to annual Australian government expenditure on fundamental spatial data of around \$70 million, it would suggest net welfare benefits from providing free access over cost recovery of around \$25 million per annum.

### Supplementary information

Drawing on a range of published sources it is possible to add a little to this overall picture. In terms of agency transaction cost savings, it has been reported that Western Australia's Landgate agency estimate their transaction and support related costs at around 17% of fundamental data revenue, the ABS (above) put savings at around 32% of revenue lost, and UK OpenData put savings at around 1/3<sup>rd</sup> of revenue lost. Scaled to average annual estimated revenue declines from scheduled datasets these proportions would suggest possible agency transaction cost savings of around \$375,000 per annum. DITR (2004) noted that in 2002-03 government agency annual sales revenue from spatial datasets was around \$1.3 million lower than 2001-02, due to making data freely available online. Using that as an approximation suggests an annual agency net loss of some \$925,000.

If user transaction and access costs and savings associated with the same transactions mirrored those of agencies, then users' transaction-related net cost savings may have been of the order of \$1.7 million per annum.

### 3.2.2 Indicators of use (web statistics)

Spatial data are very varied, with varied users and uses, and are held by a variety of agencies. At the end of September 2010, there were 21 repositories within the Australian Spatial Data Directory with a total of 28,917 dataset descriptions available for searching.<sup>51</sup> This makes it difficult to track the delivery of data, website hits and downloads. Nevertheless, from 2001-02 through 2005-06 some useful, if incomplete, data were reported in the Spatial Data Policy Executive (SPDE) Annual Reports, which can shed some light on demand elasticity. Australian Spatial Data Directory (ASDD) quarterly technical reports cover January 1999 through September 2010,<sup>52</sup> and can also shed light on demand trends.

However, there are a number of limitations and issues to consider when exploring website statistics and trying to isolate the impacts of making data freely available from other trends and influences. As noted, these include the impacts of:

- General trends in the use of online materials (*e.g.* the mid-2000s saw an increase in downloads across the board);
- Annual and cumulative publication volumes (*e.g.* there are likely to be more downloads as more is made available to download);
- Changes to other related factors (*e.g.* pricing, policy cycles, etc.); and
- Limitations in how website hits and downloads are recorded and reported, including incomplete reporting and unrecorded delivery.

In the following paragraphs we attempt to explore the extension and intensification of use as best we can in view of these issues and limitations.

While subject to a number of data-related deficiencies, the SPDE Annual Reports produced from 2001-02 through to 2005-06 provide useful information on the spatial data related activities of the major Australian government agencies, including the number of scheduled and non-scheduled datasets delivered in each format (Table 10). From 2001, online downloads were

free, standard packages were delivered at the marginal cost of transfer \$99, and customized products were delivered at the full cost of transfer.

**Table 10** Spatial datasets delivered by Australian government agencies, 2001-02 to 2005-06

	2001-02	2002-03	2003-04	2004-05	2005-06
<b>Total number of datasets delivered:</b>					
Online downloads	..	44,745	125,321	339,915	904,014
Standard packages	..	11,872	7,523	8,895	414,601
Customised products	..	33,821	213,710	362,821	205,591
Total	75,310	90,438	346,554	711,631	1,524,206
<b>Number of scheduled datasets delivered:</b>					
Online downloads	39,387	43,434	48,055	215,502	854,196
Standard packages	7,460	8,227	2,947	1,811	7,188
Customised products	28,463	31,388	1,563	2,508	1,146
Total	75,310	83,049	52,565	219,821	862,530
<b>Number of datasets available (at end of year):</b>					
Scheduled (incl. Auxiliary list)	132	172	255	427	661
Downloads per scheduled dataset available	313	262	184	505	1,292
Total delivered per available dataset	598	500	201	515	1,305

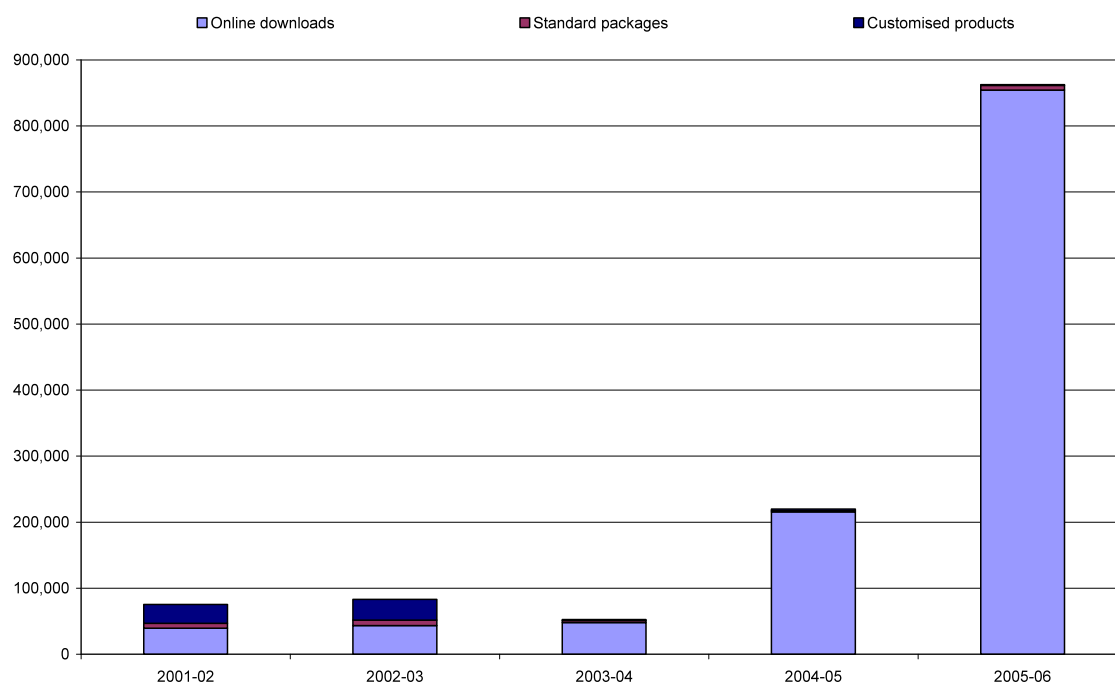
Note: The number of scheduled datasets delivered by format in 2001-02 is estimated from the shares in the following year.

Source: SPDE Annual Reports. Author's analysis.

Both the total number and number of scheduled datasets delivered increased rapidly over the period. The total number of datasets delivered increased from around 75,000 to more than 1.5 million or by 112% per annum, while the number of scheduled datasets delivered increased from around 75,000 to 863,000 or by 84% per annum. Downloads of online datasets free-of-charge increased even more rapidly: total datasets delivered online by 172% per annum and scheduled datasets delivered online by 116% per annum. These data suggest a rapid increase in the use of freely available online data.

The intensity of use also increased even as more datasets became available, with the total number of datasets delivered per scheduled dataset available increasing by 23% per annum and the number of downloads per scheduled dataset available increasing by 44% per annum (Figure 9).

**Figure 9** Scheduled spatial datasets delivered by Australian government agencies, 2001-02 to 2005-06



Source: SPDE Annual Reports. Author's analysis.

Not surprisingly, government revenue from the sale of spatial data declined, by around 4.5% per annum over the 2001-02 to 2005-06 period. With free online delivery, revenue from sales of scheduled datasets collapsed, by 22% per annum (Table 11). Hence, Australian government revenue per dataset delivered fell from around \$70 in 2001-02 to less than \$3 in 2005-06, and revenue per scheduled dataset delivered from around \$30 to \$1.

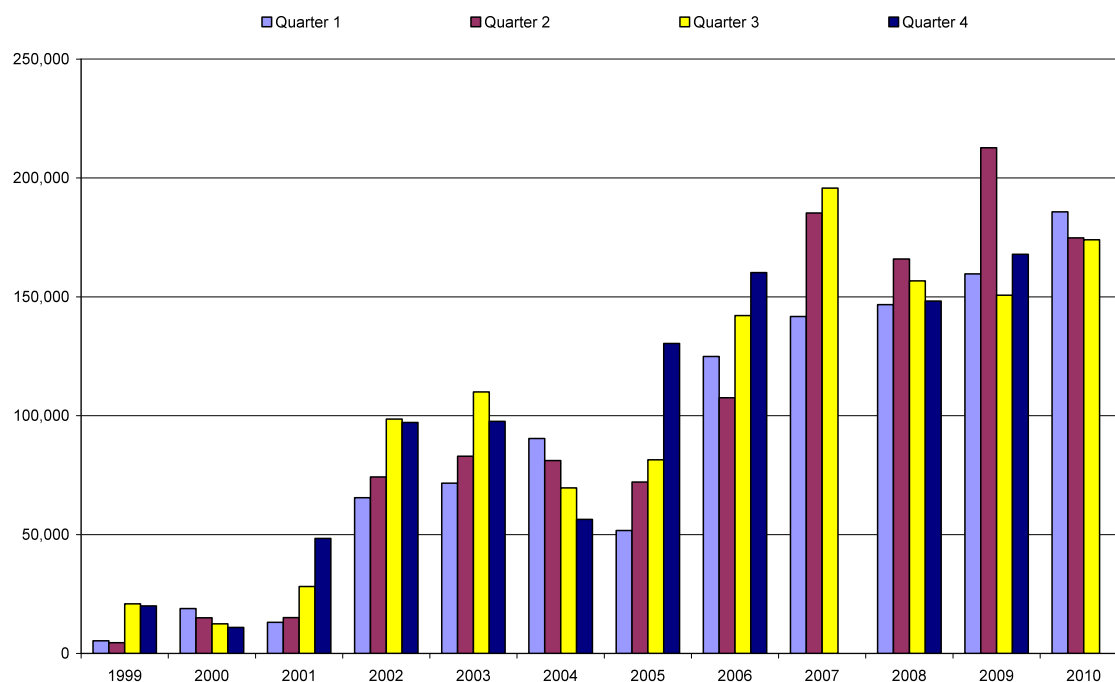
**Table 11** Australian government revenue from sales of spatial data, 2001-02 to 2005-06 (\$ millions)

	2001-02	2002-03	2003-04	2004-05	2005-06	CAGR
Scheduled	2.4	1.7	1.1	2.4	0.9	-22%
Non-scheduled	2.9	0.8	3.1	2.9	3.5	4.8%
Total	5.3	2.5	4.2	5.3	4.4	-4.5%

Source: SPDE Annual Reports. Author's analysis.

The Australian Spatial Data Directory (ASDD) reports quarterly hits, which show trends in spatial data searching and use over a longer period (*i.e.* 1999 through 2010). “Hits” refers to total page views (*i.e.* access of a page as a whole, such as html), which progressively attempt to exclude Spider page views – most notably from the beginning of 2005. Over the decade 1999 through 2009 hits increased by 30% per annum, and showed a rapid increase following the introduction of the Commonwealth Policy on Spatial Data Access and Pricing (SDAP), which was announced in September 2001 and introduced in February 2002 (Figure 10).

Figure 10 Total page views or 'hits' reported by ASDD, 1999-2010



Source: <http://asdd.ga.gov.au/asdd/tech/quarterly.html>. Author's analysis.

While no more than indicative, these data provide a basis for estimating the elasticity of demand for geospatial data.

### 3.2.3 Wider impacts of use

For estimates of the wider impacts of making fundamental spatial data more accessible we explore possible welfare impacts and returns to expenditure on geospatial science and data.

#### Consumer surplus

As noted, the period chosen as the start and end date makes a substantial difference to elasticity estimates and there are many factors influencing downloads and spatial dataset delivery trends, in addition to price and licensing changes. Nevertheless, tracking revenue per scheduled dataset delivered 2001-02 to 2005-06 suggests a price elasticity of 0.97, and tracking the total number of scheduled datasets delivered per available scheduled dataset over the same period suggests a demand elasticity of 1.3 (Table 15). Hence, the price elasticity of demand for scheduled datasets over the period was 1.3.

Revenue trends are also difficult to interpret. Between 2001-02 and 2002-03, Australian government revenue from sales of spatial data fell by \$2.8 million, of which around \$700,000 was from scheduled datasets. Over the period 2001-02 to 2005-06 revenue from sales of scheduled datasets fell by \$1.5 million per annum, and the difference between 2001-02 revenue and the annual average for the subsequent years was \$875,000 per annum for scheduled datasets

and \$1.2 million for all datasets. Hence, the range for revenue lost might be of the order of \$700,000 to \$1.5 million per annum. DITR (2004) reported a revenue loss by Geoscience Australia of \$1.3 million per annum, which would suggest an annual increase in consumer surplus of around \$2.2 million. Taking average annual government expenditure on fundamental data over the period 2001-02 through 2005-06 as indicative of the costs that would have been recovered under a cost-recovery regime, suggests an annual consumer surplus increase of around \$60 million (Table 12).

**Table 12** Estimates of annual consumer surplus and increase in returns to annual spending

	<i>Elasticity / Value</i>
<b>Elasticity estimates:</b>	
Revenue/scheduled dataset 2001-02 to 2005-06	0.97
Revenue/total dataset 2001-02 to 2005-06	0.96
Online downloads per scheduled dataset available	3.3
Total delivered per scheduled dataset available	1.3
<i>Price elasticity of demand for scheduled datasets</i>	1.3
<b>Welfare estimates:</b>	
GA annual revenue foregone 2001-02 to 2005-06	1,300,000
Consumer surplus	2,164,970
At cost recovery (government spending on scheduled fundamental data)	36,174,465
Consumer surplus	60,234,460
<b>Returns estimates:</b>	
Average annual expenditure on fundamental data	88,139,657
Share that is scheduled data	36,174,465
At 20% return (assuming the average from R&D)	7,234,893
Increase in total delivered per dataset available 2001-02 to 2005-06	129%
Increase in returns (from downloads per page published 3yr 2003-04 to 2005-06 and 2006-07 to 2008-09)	15,496,191
Note: Elasticities are expressed as absolute values.	
Source: SPDE Annual Reports. Author's analysis.	

These estimates do not include any deadweight loss associated with compensating for the loss in revenue through additional government funding, as Geoscience Australia did not seek additional funding. Nor do they include producer surplus as the data are collected for public purposes anyway and there would be no profit in an alternative cost-recovery regime. It is also worth noting that government is a major user of spatial data.

### Returns to expenditure

To the extent that there are similarities between spatial science and fundamental spatial data and the data arising from observational research, it is possible to use methods relating to returns to R&D to explore the wider impacts (See Box 3). For the reasons noted, as a first approximation, we assume a lower bound 20% return on expenditure. Returns to research expenditure are

recurring over the useful life of the knowledge, but lagged to account for the time between the conduct of research and its use impacts, and discounted accordingly. Our consultations suggested that spatial data have a long useful life, with impacts distributed over a considerable period. Hence, we estimate increases in returns to a single year's expenditure over 25 years, discounted.

The impacts of the observed increases in use, as indicated by trends in scheduled datasets available and delivered over the period 2001-02 through 2005-06, on average social returns to annual expenditure are estimated as described above (Box 3). These estimates suggest an increase in social returns to annual expenditure of some \$15 million (Table 15).

### 3.2.4 Summary of impacts

Approximate estimates suggest possible agency savings on transactions of around \$375,000 per annum and revenue losses of \$1.3 million per annum. Conversely, users' savings include both (approximately \$1.7 million per annum). Pwc (2010) computed welfare impacts of free provision of GA topographical data relative to cost recovery and noted an overall increase in net welfare of \$4.7 million per annum. The impacts of the observed increases in use, as indicated by trends in scheduled datasets available and delivered over the period 2001-02 through 2005-06, on average social returns to annual expenditure on data collection suggest an increase in social returns of some \$15 million. Hence, the benefits may be around 13 times the costs in terms of revenue foregone (Table 13).

**Table 13 Summary of spatial data case study annual costs and benefits (circa 2005-06)**

<i>Item</i>	<i>Value</i>
<b>Agency and user impacts:</b>	
Agency impacts (net)	-923,900
Users impacts (net)	1,676,100
<b>Welfare estimates:</b>	
Surplus from lost revenue (2001-02 to 2005-06)	2,165,000
Surplus from cost recovery (scheduled fundamental data)	60,243,500
<b>Returns estimates:</b>	
Increase in returns (downloads per page published 3yr 2003-04 to 2005-06 and 2006-07 to 2008-09)	15,496,200
<b>Overall impacts estimates:</b>	
Estimated total costs (revenue foregone)	1,300,000
Estimated total benefits (savings and returns)	17,548,400
Benefit/cost ratio	13
Source: Various sources. Author's analysis.	

While no more than exploratory, these estimates might be seen to provide approximate estimates of lower and higher bound impacts of making fundamental spatial data freely available online.

### 3.3 Hydrological Data (National Water Commission & Bureau of Meteorology)

In 2005, the National Water Commission (NWC) produced a baseline assessment of water resources in Australia, *Australian Water Resources 2005* (AWR 2005).<sup>53</sup> The primary purpose of AWR 2005 was to provide a baseline picture of a range of water management and resource issues from which future comparisons and the success of National Water Initiative (NWI) reform processes could be measured. It focused on water availability, water use and river and wetland health. It also identified water resource knowledge gaps critical to the success of NWI measures, and provided a framework for the establishment of an ongoing water data information infrastructure: the Australian Water Resources Information System (AWRIS).

AWR 2005 identified the lack of an agreed framework and data to undertake a national river and wetland health assessment, and noted a number of gaps in the consistency, extent and availability of information to provide for robust resource management, including:

- An inability to provide information in a readily accessible manner across states and territories;
- A lack of standard terminology and definitions for key parameters, including sustainable yield and over allocation of water resources; and
- A lack of integration between those organizations managing water availability (resource allocation) and those managing the health of high value environmental assets (river and wetland health).

The information collected to that time was based on separate standards and subject to different licensing requirements. Of the total \$3.5 million spent on the 2005 assessment, it is estimated that around 10% (\$350,000) was spent on data collection issues.

*The Water Act 2007* brought an extension of responsibilities for the Bureau of Meteorology (BOM), seeing it take over the management of water information and the establishment of the Australian Water Resources Information System (AWRIS). These functions were supported by the Water Regulations of 2008, which came into effect on 30 June that year. The Regulations define who must give specified water information to the Bureau, and the time and format in which it must be supplied.<sup>54</sup> The Regulations name more than 200 organizations that are required to give the BOM specified water information that is in their possession, custody or control.

The BOM now publishes an annual *National Water Account* (NWA), which reports on the total water resource, the volume of water available for abstraction, the rights to abstract water, and the actual abstraction of water for economic, social, cultural and environmental purposes across Australia. The NWA provides information that has previously been difficult to access or unavailable to general users in a standardised form. It enables national comparability of



information and highlights gaps and inconsistencies in data and knowledge, allowing improvements to be made. The information presented in the NWA strives to be: nationally consistent and comparable; publicly available; comprehensive; useful to governments, the water sector, industry and the general community; and transparent about the source and quality of the data.<sup>55</sup>

In order to make the data available and useful, BOM is making it freely available online and has adopted Creative Commons licensing. In late 2010, BOM reported that they had secured the commitment of 182 out of 215 water data providers (85%) to apply CC licensing to the data they supply to them,<sup>56</sup> and at the time of writing just nine data providers had not yet agreed to CC licensing.

### 3.3.1 Agency and user costs and benefits

Water data has been fragmented and it is too early to measure the impacts of making National Water Account data available centrally. Consequently, a detailed case study is not possible.

#### Victorian Water Resources Data Warehouse

However, the *Victorian Water Resources Data Warehouse* provides an example of the cost and use impacts of making water data freely available online. Secondary sources and informal consultations suggest that annual water data collection costs in Victoria were running at around \$6.7 million in the late 1990s, of which 60% or \$4 million related to water quantity data, 25% or \$1.7 million to ground water data and 15% to water quality data. At that time, there were around 600 requests for water quantity data and 100 for water quality data per annum, suggesting an average cost per user request of some \$7,180. It was reported that only around 5% of requests were from the private sector.<sup>57</sup>

Immediately prior to the establishment of the *Victorian Water Resources Data Warehouse* in 2004, which disseminates up-to-date information on Victoria's water resources through the World Wide Web that gives access to both raw and summary data on water quality and quantity throughout Victoria, there were three contractors serving user data requests at a cost of around \$300,000 per annum. They handled around 400 requests per annum, suggesting an implied cost of around \$750 per user request. Establishing the *Victorian Water Resources Data Warehouse* was a 5-year project costing around \$870,000. In the early years of its operation it was reported to have handled 60,000 data requests in the first six months, at an implied cost of \$1.45 each.<sup>58</sup> At the time of writing, the site reported a total of 1,257,958 visitors (an average of approximately 180,000 per annum). While not all visits amount to a data request, it seems that use increased and costs per use have fallen dramatically.

#### Bureau of Meteorology

By the end of 2010, the *Bureau of Meteorology* (BOM) had received more than 8 million water data files from over 200 agencies across the nation, and was receiving more than 10,000 new files per day. The data encompasses 75 variables across 10 categories of water information, including streamflow, groundwater, climate, water storages and water entitlements, allocations, trades and restrictions.<sup>59</sup> While there are no data on the transaction costs involved or a

comparable prior system, it is clear that agency and data provider costs are being kept to a minimum through clear delineation of responsibilities, standardisation of data formats and licensing conditions.

On the user side, the Bureau is striving to make most of Australia's water information freely and publicly accessible, and to package it in a way that maximises its utility. This will not only assist water managers and policy makers to do their jobs better. It will also help to satisfy the needs of water-dependent businesses, farmers, industry, educators and the general community.

### 3.3.2 Indicators of use (web statistics)

As noted, water data has been fragmented and it is too early to measure the impacts of making National Water Account data available. Nevertheless, web statistics from the *Victorian Water Resources Data Warehouse* and the *National Water Commission* give some indication of usage trends when water data are made freely available online.

#### Box 4 The Victorian Water Resources Data Warehouse

The *Victorian Water Resources Data Warehouse* gives access to both raw and summary data on water quality and quantity throughout Victoria, and is a central repository for published documents produced from this data. The site offers you a number of choices in how you access this information:

*Maps for Site Selection* – allow you to narrow down the sites you are interested in by clicking on a map of the State. Once you have chosen your site in this way, the relevant data will be accessible.

*Standard Warehouse Reports* – provides access to summary information for sites such as average monthly flows, annual summary statistics and active site lists.

*Individual Site Information* – use this link if you already know the site name or number you are interested in and want to find out information about the site such as parameters measured, location, contractor, rating table etc.

*Extract Data from Warehouse* – allows you to develop a query to narrow down the data you are interested in. For example you could extract a site list for sites that are both in the West Gippsland CMA, and that have phosphorus results greater than 1 mg/l. There is a range of parameters which you can alter to adjust your query.

*Browse Statistics and Data* – allows you to browse through summary statistics and "drill down" into the data. For example you might look at annual flows for a range of sites and then choose to look at the monthly flows for a particular site by clicking on the annual flow result.

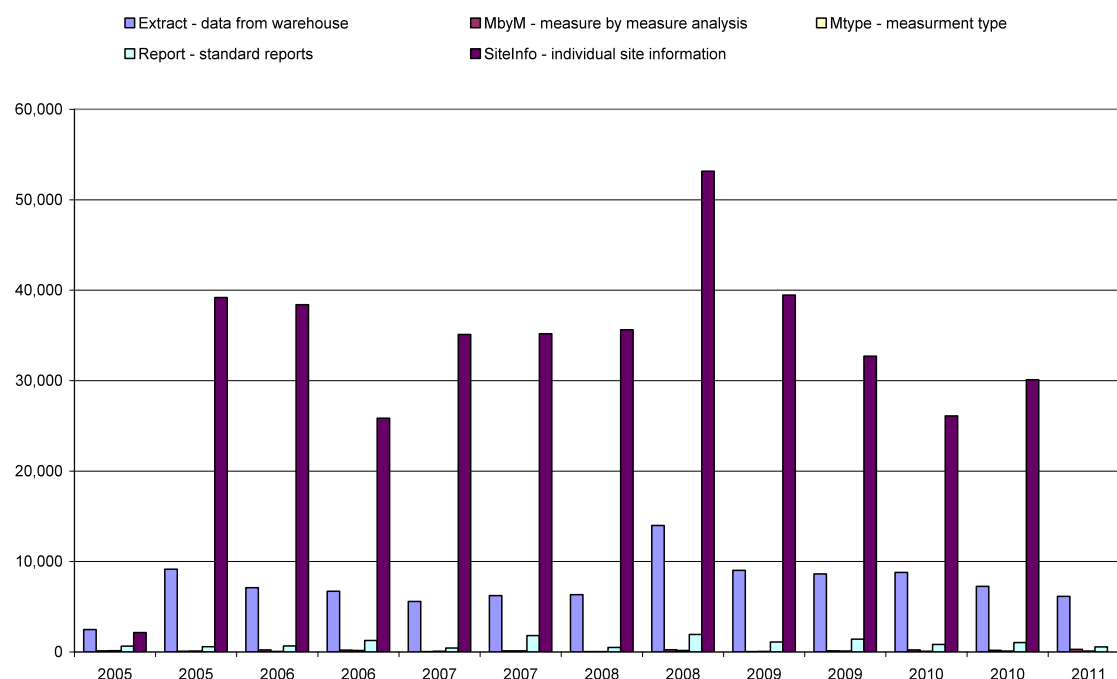
*Measure by Measure Analysis* - is similar to the Extract Data from Warehouse page but it allows you to create a graph of one measure versus another. For example you could look at stream flow versus total phosphorus results for a site or examine both parameters as a time series.

*DSE Published Documents* – provides access to documents published as part of the water resources monitoring program of Department of Sustainability and Environment, including trend analysis for the whole state.

Source: <http://www.vicwaterdata.net/vicwaterdata/home.aspx>

Monthly data from the *Victorian Water Resources Data Warehouse* show growth in use and re-use, with online requests for information about water conditions at individual sites increasing from 2,155 per month in April 2005 to more than 35,000 per month by late 2007, and data extractions (*i.e.* extracting data for re-use) increasing from 2,476 per month in April 2005 to almost 14,000 by late 2008 (Table 14 and Figure 11).

**Figure 11** Data use and re-use from the Victorian Water Resources Data Warehouse (monthly, April and October in each year)



Source: The Victorian Water Resources Data Warehouse. Author's analysis.

Prior to the establishment of the *Victorian Water Resources Data Warehouse* data requests were reported to be running at around 400 per annum. The free availability of the data has had a marked impact on use:

- Use of *individual site information* increased from 185,000 in 2005 to more than 360,000 during 2010, and there have been more than 3 million site information requests to date, an average of approximately 443,000 per annum;
- *Data extractions* (*i.e.* for re-use) also approximately doubled, increasing from 68,908 during 2005 to 102,536 during 2010, and there have been almost 600,000 data extractions from the warehouse to date, an average of approximately 85,000 per annum; and
- More than 19,400 unique IP addresses have been recorded performing data extractions, and almost 22,600 IP addresses have been recorded sourcing individual site information.

**Table 14** Data use and re-use from the Victorian Water Resources Data Warehouse (monthly)

	APR 2005	OCT 2005	APR 2006	OCT 2006	APR 2007	OCT 2007	APR 2008	OCT 2008	APR 2009	OCT 2009	APR 2010	OCT 2010
Data Extraction	2,476	9,155	7,113	6,705	5,584	6,234	6,344	13,987	9,020	8,632	8,793	7,262
Measure analysis	122	83	229	204	46	128	35	255	46	126	234	184
Measurement type	145	110	71	165	83	123	51	172	63	102	74	106
Standard reports	636	594	671	1,280	442	1,814	504	1,932	1,098	1,427	840	1,044
Individual site info	2,155	39,194	38,412	25,858	35,104	35,190	35,629	53,157	39,455	32,710	26,106	30,112

Source: The Victorian Water Resources Data Warehouse. Author's analysis.

Similarly high levels of use have been recorded at the National Water Commission website (<http://water.gov.au>), where there were 336,720 visits recorded between 23 May 2005 and 22 June 2011, with 269,002 unique visitors, and a total of 796,074 page views. There have been an average of 230 visits to the site per day. Of the visitors, 80% are new and 20% returning visitors. The visits have come from 206 countries and territories, 257,752 from Australia, 23,615 from the United States, and between 7,000 and 5,000 from the United Kingdom, India and Canada. Visitors have viewed an average of 2.36 pages per visit. Reflecting the importance of discoverability, just 10.5% of visits were direct to the site, 18% came from referring sites and 72% came from search engines.

While no more than indicative, these examples suggest that water data are used extensively when made readily and freely available online, suggesting that there is a much higher return on investment in data collection when that data is made freely available.

### 3.3.3 Wider impacts of use

The centralised and coordinated management and open availability of water data is too recent a development for its impacts to be evident. However, a number of impacts are reported in general terms in the *Improving Water Information Program: Progress Report 2010*.<sup>60</sup>

*As competition for water resources intensifies, it is more important than ever to account for how water is distributed across Australia in a transparent and rigorous way. One important element is resource allocation through the National Water Markets System (NWMS), which helps to inform water markets by providing regularly updated online reports on water entitlements, allocations, trade volumes and trade prices. Improving this information should help to increase the equity of the Australian water market, valued at \$3 billion in 2009-10 – having doubled in value since 2007-08.*

*The protracted drought over many parts of Australia increased demand for more timely and accurate water availability forecasts, prompting the Bureau [BOM] to expand its services to include continuous short-term streamflow and seasonal streamflow forecasts, which are expected to have a major impact on Australian water policy development and practice. The forecasts will provide a much stronger and more timely information base for decisions about water allocations, cropping strategies, water demand, purchasing environmental water, water trading and drought management.*

*Accurate estimates of the intensity/frequency/duration characteristics of rainfall are critical to the task of designing structures affected by flooding, such as gutters, culverts, drains and bridges. These estimates are also a critical input to hydrologic models used for the assessment of flood risk. Current estimates of design rainfall published in the Australian engineers handbook Australian Rainfall and Runoff are based on data available up to 1983. Since then the availability of additional rainfall data has increased markedly and new techniques in frequency analysis have been developed.*

These examples demonstrate the importance of water information as a basis for planning and investment decisions, design and construction, crop planting and harvesting, etc.

## Box 5 Australia's Water Market

The Australian water market is a composite of many separate markets, each defined by water system boundaries and administrative arrangements. Within those markets are segments for different water products (water access entitlements, water allocations and other forms of entitlement) and different trading transactions (transfers, leases, amalgamations and so on).

In 2009-10, the total trade of entitlements in Australia was 1,949 GL, an 8% increase from 2008-09. The total Australia-wide volume of water allocation trade also increased, but by a comparatively greater margin, from 2,158 GL in 2008-09 to 2,495 GL in 2009-10 (an increase of 16%).

The total value of turnover of entitlements (including, but not limited to water access entitlements) is estimated to have been about \$2.6 billion in the 2009-10 water year, which was a 17% increase on 2008-09. The higher market value was due to an increase in the volume of entitlements traded.

Total sales of water allocations were valued at \$366 million (excluding Queensland seasonal water assignments as price data is not collected for those trades). This was a 40% decline in the estimated value of the water allocation market from 2008-09. While the volume traded rose by 16% in 2009-10, the average price was significantly lower than in the previous year.

Overall, Australia's water markets continued to increase in value, from \$2.8 billion in 2008-09 to \$3.0 billion in 2009-10.

Source: NWC (2010) *Australian Water Markets Report 2009-10*, Canberra.

The emergence of a market for water entitlements and allocations also makes a major contribution, enabling us to step back from unsustainable over-allocation and make more efficient use of a scarce resource.

## 4 Wider impacts of open access to PSI

The wider impacts of making PSI freely available include: the direct private cost-benefits of new users and uses, and more intensive use by existing users; the privately captured spillover cost-benefits of additional use for other businesses and agencies; and the public spillover cost-benefits of additional use. These can arise in the form of new activities, businesses and industries (*e.g.* Weather Derivatives, Geospatial Services, etc.); increased efficiency of existing activities, businesses and industries (*e.g.* optimisation of crop planting and harvesting, mining exploration and extraction, etc.); and public good aspects (*e.g.* cheaper food, enhanced safety while travelling, etc.). Such impacts are very difficult to measure.

As noted, there have been various approaches used to explore the value and impacts of PSI. Some of the earlier top-down approaches resulted in high estimates of the value of PSI, but did not account for demand factors (*e.g.* the existence of substitutes) or their impacts on potential multiplier effects. Bottom-up approaches have typically explored net surplus or welfare impacts and estimated the value of the PSI-related consumer surplus, but are vulnerable to data limitations and assumptions about demand. For example, the elasticity of demand can be affected by a number of factors, including:

- The type of information being considered (*e.g.* for highly valuable and time-critical safety and warning information demand may be inelastic, whereas for historic map information demand may be more elastic);
- The proximity of substitutes (*e.g.* where there are close substitutes demand may be more elastic as users can switch between substitute products in response to price changes, whereas basic information like company registration information may have no close substitutes); and
- Where the information is an input to value-adding, the share of input costs accounted for by the information will be important (*e.g.* the smaller the proportion of total cost of the refined information product that can be attributed to an input, the more inelastic the demand for such an input will be because the incentives to switch to alternatives are relatively weak when the PSI constitutes a small proportion of the overall cost base and the impact of an increase in the cost of PSI on final good prices is relatively small).<sup>61</sup>

In Australia, both top-down and bottom-up approaches have been used to explore the impacts of making fundamental spatial data freely available. In this section, we look at the reported impacts of spatial data in Australia and then at the impacts of PSI more generally.

### 4.1 Reported impacts of spatial data in Australia

Pwc (2010) computed the welfare impacts of the free provision of GA topographical data relative to cost recovery and noted: an agency loss of \$13.3 million per annum in revenue foregone; a gain for government users of \$10 million per annum; a gain for private consumers of \$8 million per annum; and an overall increase in net welfare of \$4.7 million per annum.

Extrapolating this to annual Australian government expenditure on fundamental spatial data of around \$70 million would suggest net welfare benefits of around \$25 million per annum.

Looking at spatial information in Australia, ACIL Tasman (2008) estimated that industry revenue in 2006-07 could have been of the order of \$1.37 billion and industry gross value added around \$682 million. Using a General Equilibrium (GE) modelling approach, they concluded that the economic footprint of the spatial information industry was larger.<sup>62</sup>

To support the recent review of Geoscience Australia, the Department of Finance and Deregulation commissioned ACIL Tasman to report on the economic value of the core areas of GA's work, including that relating to pre-competitive geological information on petroleum and minerals, and its work in gathering, processing and disseminating geospatial, Earth monitoring and groundwater information.<sup>63</sup> In both areas, ACIL Tasman investigated the public good aspects of the products and services produced by GA, and the private and public benefits flowing from them.

ACIL Tasman divided impacts into productivity and non-productivity benefits. Productivity benefits are realised in areas such as logistics, precision agriculture, asset mapping, and infrastructure maintenance. Non-productivity benefits arise through improving public decision making, natural resource management including water management, natural disaster and emergency management and national security.

Looking at productivity benefits, ACIL Tasman adopted a representative figure of \$12 billion in productivity benefits for 2010. They considered that it would be conservative to attribute 15% of the overall productivity benefits to GA's efforts. This results in an estimated increase in GDP due to the accumulated impact of GA's provision of geospatial products and services of \$1.8 billion for 2010.

In terms of non-productivity benefits, ACIL Tasman examined the benefits associated with improvement in public decision-making underpinned by GA's products and services in the areas of natural resource and environmental management, water management, natural disasters and emergency management, and national security. ACIL Tasman estimated:

- The annual benefits accruing from the use of geospatial information in support of applications, such as managing dryland salinity, monitoring and managing the impacts of climate change, water quality management, locust control and biosecurity amount to \$1.1 billion;
- The annual water management benefits from improved water-use efficiency and improved investment in water infrastructure routing to be \$100 million; and
- The annual benefits associated with mitigating natural disasters in areas where GA is involved, such as tsunami warning, bushfire management and flood control to be \$500 million.

Overall, ACIL Tasman considered that a plausible estimate of the non-productivity benefits of geospatial, Earth monitoring, groundwater and hazards information is \$1.7 billion per annum. They did not attempted to assign a particular percentage of this benefit to GA, given the



difficulty in attributing value between various fundamental data acquisition and value adding activities. However, they claimed that a large portion was likely to be attributable to GA.

## 4.2 Impacts of PSI more broadly

Key findings from some of the major international studies looking at the value and impacts of PSI were noted above. Unfortunately, no such study has been conducted in Australia, and we can only extrapolate from the international evidence.

The influential study by PIRA (2000) estimated the investment value of PSI (*i.e.* what governments invest in the acquisition of PSI) and the economic value of PSI (*i.e.* the national income attributable to activities built on the exploitation of PSI). In the European Union, they put the former at around EUR 9.5 billion per annum in 1999 and the latter at around EUR 68 billion (equivalent to approximately 1.4% of EU GDP).<sup>64</sup> Economic structures vary from country to country, but assuming similar levels of investment and use in Australia, this would be equivalent to an investment value for PSI of the order of \$2.5 billion and a use value of around \$18 billion in today's prices.

In the MEPSIR study, Dekkers *et al.* (2006) put the overall market for PSI in the EU plus Norway at around EUR 27 billion (approximately 0.25% of aggregated GDP).<sup>65</sup> Assuming similar levels of activity in Australia, this would be equivalent to \$3.2 billion in today's prices.

Making some adjustments to the MEPSIR estimates with the benefit of hindsight, te Velde (2009) suggested that the value might drop further from EUR 27 to EUR 5 billion or even EUR 3 billion.<sup>66</sup> Again, assuming similar levels of activity in Australia, this would reduce the value of the PSI market to around \$500 million.

In their report to the UK Office of Fair Trading, DotEcon (2006) suggested that the net value of PSI in the UK was around GBP 590 million per annum in 2005.<sup>67</sup> Assuming similar levels of activity in Australia, this would suggest a PSI value of the order of \$2.4 billion in today's prices.

While there are many studies of the value of PSI of various types, what these examples suggest is that: *(i)* the value of PSI is substantial, but very difficult to estimate with certainty; and *(ii)* the use value of PSI depends upon accessibility and usability.



## 5 Lessons for the research sector

Research publications and data differ from other forms of PSI and are not always included in discussion and analysis of PSI. For example, the OECD and US National Academies both include publicly funded research publications and data in PSI, but it is subject to separate analysis, policies and regulations. Open access policies for research publications in the form of research articles are well developed, and an increasing number of funding bodies and research institutions are now pushing for open access to research data. The OECD has implemented a declaration on access to research data from public funding (2007)<sup>68</sup> and issued a recommendation on Public Sector Information (PSI) (2008),<sup>69</sup> as has the European Commission and a number of individual countries.

### Box 6 Benefits of curating and sharing publicly funded research data

Open access to, and sharing of, data from publicly funded research offer many research and educational advantages over a closed, proprietary system that places high barriers to both access and subsequent re-use. Open access to such data:

- Reinforces open scientific and scholarly inquiry;
- Encourages diversity of analysis and opinion;
- Promotes new research and new types of research;
- Enables the application of automated knowledge discovery tools online;
- Allows the verification of previous results;
- Makes possible the testing of new or alternative hypotheses and methods of analysis;
- Establishes a broader base set of data than any one researcher can hope to collect, thereby providing a greater baseline of factual information for the research community;
- Supports studies on data collection methods and measurement;
- Facilitates the education of new researchers;
- Enables the exploration of topics not envisioned by the initial investigators;
- Permits the creation of new data sets, information, and knowledge when data from multiple sources are combined;
- Helps transfer factual information to, and promote development and capacity building in developing countries;
- Promotes interdisciplinary, inter-sectoral, inter-institutional, and international research; and
- Helps to maximize the research potential of new digital technologies and networks, thereby providing greater returns from the public investment in data collection and research.

Source: GEO (2010) *GEOSS Data Sharing Action Plan*. GEO-VII Plenary.

The same basic economic principles apply to publicly funded research outputs and PSI, namely:

- Governments around the world fund research and the creation of public sector information because it has public good characteristics (*i.e.* is more or less non-rivalrous in consumption and it is difficult to enforce exclusion);
- The Internet has drastically reduced the cost of distribution of information (*i.e.* the marginal cost of distribution of one additional copy is close to zero); and
- The efficient price for distribution of public goods is no more than the marginal cost of distribution (*i.e.* effectively zero). Any other price leads to the inefficiency of excluding additional users who could have obtained access at the marginal cost of distribution (*i.e.* zero).

However, methods for measuring the impacts of enhanced access to PSI and research outputs often differ. Much more is known about the impacts of publicly funded research and returns to R&D expenditure than is known about the impacts of PSI. Conversely, some forms of PSI have been priced at cost recovery, whereas publicly funded research has not. This affects preferred approaches to valuing the information and/or valuing increases in access to it. While the use of case studies and extrapolation is common to both, it is difficult to adopt a welfare economics approach where access has characteristically been free, and difficult to adopt a return to investment approach where returns have rarely been traced or measured.


## 5.1 Cost-benefit analyses of research data curation and sharing

Efforts to understand the costs and benefits involved in research data curation and sharing are more advanced than is the case with PSI, as a number of studies have addressed the topic in detail. They typically mix quantitative and qualitative methods, and rely primarily on case studies and extrapolation there from. Some have provided templates for assessing costs and benefits.

For example, the series of projects under the title Keeping Research Data Safe (KRDS) has explored the costs and benefits of research data curation and sharing in the UK and elsewhere in Europe. The initial KRDS study (Beagrie *et al.* 2008) investigated the medium to long-term costs to Higher Education Institutions (HEIs) of the preservation of research data, and provided a brief overview of the potential benefits to HEIs from preservation of research data.<sup>70</sup> It developed a framework and guidance for determining costs consisting of: a list of key cost variables and potential units of record; an activity model divided into pre-archive, archive, and support services and divided into the major phases from an activity model and by duration of activity; and a resources template including major cost categories. A series of case studies from Cambridge University, King's College London, Southampton University, and the Archaeology Data Service at York University, illustrated different aspects of costs for research data within HEIs. Selective illustrations of cost-benefits and costs over time were also provided.<sup>71</sup> Importantly, the study noted that the costs of a central data repository are an order of magnitude greater than that suggested for a typical institutional repository focused on e-publications alone.

Figure 12 The KRDS2 benefits taxonomy

Dimension 1	
Direct Benefits	Indirect Benefits (Costs Avoided)
New research opportunities	No re-creation of data
Scholarly communication/access to data	No loss of future research opportunities
Re-purposing and re-use of data	Lower future preservation costs
Increasing research productivity	Re-purposing data for new audiences
Stimulating new networks/collaborations	Re-purposing methodologies
Knowledge transfer to industry	Use by new audiences
Skills base	Protecting returns on earlier investments
Increasing productivity/economic growth	
Verification of research/research integrity	
Fulfilling mandate(s)	

Dimension 2	
Near Term Benefits	Long-Term Benefits
Value to current researcher & students	 Secures value to future researchers & students.  Adds value over time as collection grows and develops critical mass
No data lost from Post Doc turnover	
Short-term re-use of well curated data	
Secure storage for data intensive research	
Availability of data underpinning journal articles	

Dimension 3	
Private Benefits	Public Benefits
Benefits to sponsor /funder of research/archive	Input for future research
Benefits to researcher	Motivating new research
Fulfil grant obligations	Catalysing new companies and high skills employment
Increased visibility/citation	
Commercialising research	

Source: Beagrie, N., Lavoie, B, and Wollard, M. (2010) *Keeping Research Data Safe 2*, JISC, London and Bristol, p55. Available <http://www.beagrie.com/jisc.php> (April 2011).

A second phase project, Keeping Research Data Safe 2 (KRDS2), further developed the activity-based cost model presented in the original study, presented detailed cost information for four organizations (the Archaeology Data Service, National Digital Archive of Datasets, UK Data Archive, and University of Oxford), and developed a benefits framework illustrated with two benefit case studies from the National Crystallography Service at Southampton University and the UK Data Archive at the University of Essex. The study found that: the costs of archiving activities are a very small proportion of the overall costs and significantly lower than the costs of acquisition/ingest or access activities; and there can be significant benefits in the short-term to current researchers, as well as long-term benefits to future research.<sup>72</sup>

Fry *et al.* (2008) sought to identify benefits arising from the curation and open sharing of research data.<sup>73</sup> They suggested that potential benefits include: maximised return on investment in data collection; broader access where costs would be prohibitive for individual researchers/institutions; potential for new discoveries from existing data, especially where data are aggregated and integrated; reduced duplication of data collection costs and increased transparency of the scientific record; increased research impact and reduced time-lag in realising those impacts; new collaborations and new knowledge-based industries. Broader, indirect benefits might include transparency in research and funding, use of data sets in education to enhance the data awareness of students, enhanced researchers' skills through access to a broader range of data, tools and standards having the potential to increase data quality, and increased visibility and promotion of institutions and researchers.<sup>74</sup>

The Fry *et al.* study used a mixed-method approach, including a literature review and qualitative case studies to inform the development of a model on which to build a business case for data sharing in UK Higher Education. This was based on extensions to the research data preservation cost model proposed by Beagrie *et al.*, to allow estimation of cost-benefits to users depositing or accessing data. The case studies investigated were the European Bioinformatics Institute (EBI) and Qualidata, which is part of the Economic and Social Data Service. Based on the work of co-authors Houghton and Rasmussen, the report presented a simple example of cost-benefit analysis applicable to an individual dataset or repository, based on costs and potential cost savings. It described the data requirements and walked the reader through the process step-by-step. The approach was then extended to explore the more diffuse benefits of data curation and sharing at the institutional and disciplinary levels. Importantly, the report included an outline questionnaire and template to facilitate cost-benefit analysis.<sup>75</sup>

These and ongoing projects provide guidelines and templates that could be employed in the analysis of the costs and benefits of research data curation and sharing, and might be adapted for application in the analysis of other forms of PSI.

## **5.2 Lessons for the research sector and next steps**

The publications and data arising from publicly funded research differ somewhat from other forms of PSI. Consequently, it is difficult to draw direct lessons for the research sector from the case studies explored in this study. Nevertheless, it is clear that many of the same issues arise when attempting to measure the value of the data and/or the costs and benefits associated with providing open access to it.

The evidence from previous studies suggests that individual cases vary greatly, making generalisation extremely difficult and of limited value. Perhaps, what could more usefully be generalised are the methods of analysis. For example, it would be useful to combine the frameworks and models into a tool that could be applied in assessing the costs and benefits of research data curation and sharing, and to further develop the framework for estimating cost-benefits outlined in this study to produce a tool tailored to the analysis of the costs and benefits of providing open access to PSI. These tools might consist of a template for data collection, a draft questionnaire outlining the questions required to elicit the necessary information, and a simple spreadsheet-based online model that people could use to perform a cost-benefit analysis. The models should include all quantifiable costs and benefits, but must also include qualitative issues to help to prioritise data preservation, access and curation projects (*e.g.* incorporate a balanced scorecard approach to weighing the more intangible benefits).

Studies of the costs and benefits associated with open access to publications and data arising from publicly funded research suggest that the benefits can be significant and the costs relatively small. The return on investment in curation and open accessibility is all the greater when the data are long-lived, as the returns are recurring during the useful life of the data, although the time lag between investment and return (cost and benefit) can be substantial. What this study demonstrates is that the direct and measurable benefits of making PSI available freely and unrestrictedly typically outweigh the costs. When one adds the longer-term benefits that we cannot fully measure, cannot even foresee, the case for open access appears to be strong.

---

## Endnotes and references

- <sup>1</sup> Nilsen, K. (2007) *Economic Theory as it Applies to Statistics Canada: A Review of the Literature*, Statistics Canada, Ottawa. Available [http://www.epsiplus.net/media/files/nilsen\\_economics\\_paper\\_2007\\_final\\_version](http://www.epsiplus.net/media/files/nilsen_economics_paper_2007_final_version) (April 2011); and Nilsen, K. (2009) 'Enhancing Access to Government Information: Economic Theory as it Applies to Statistics Canada,' in Paul Uhler (ed.) *The socioeconomic effects of public sector information on digital networks: Towards a better understanding of different access and reuse policies*, National Academies Press, Washington DC. Available [http://www.nap.edu/openbook.php?record\\_id=12687&page=40](http://www.nap.edu/openbook.php?record_id=12687&page=40) (April 2011).
- <sup>2</sup> Stiglitz, J.E., Orszag, R.R and Orszag, J.M. (2000) *The Role of Government in a Digital Age*, Computer and Communications Industry Association, Washington, DC. Available <http://www.ist-world.org/ResultPublicationDetails.aspx?ResultPublicationId=5b3ae433c24747f489e0ad4470aa810b> (April 2011).
- <sup>3</sup> Nilsen, K. (2007) *Economic Theory as it Applies to Statistics Canada: A Review of the Literature*, Statistics Canada, Ottawa. Available [http://www.epsiplus.net/media/files/nilsen\\_economics\\_paper\\_2007\\_final\\_version](http://www.epsiplus.net/media/files/nilsen_economics_paper_2007_final_version) (April 2011); and Nilsen, K. (2009) 'Enhancing Access to Government Information: Economic Theory as it Applies to Statistics Canada,' in Paul Uhler (ed.) *The socioeconomic effects of public sector information on digital networks: Towards a better understanding of different access and reuse policies*, National Academies Press, Washington DC. Available [http://www.nap.edu/openbook.php?record\\_id=12687&page=40](http://www.nap.edu/openbook.php?record_id=12687&page=40) (April 2011).
- <sup>4</sup> Pollack, R. (2009) *The Economics of Public Sector Information*, CWPE 0920, University of Cambridge, Cambridge. Available <http://www.econ.cam.ac.uk/dae/repec/cam/pdf/cwpe0920.pdf> (April 2011).
- <sup>5</sup> Nilsen, K. (2009) 'Enhancing Access to Government Information: Economic Theory as it Applies to Statistics Canada,' in Paul Uhler (ed.) *The socioeconomic effects of public sector information on digital networks: Towards a better understanding of different access and reuse policies*, National Academies Press, Washington DC. Available [http://www.nap.edu/openbook.php?record\\_id=12687&page=40](http://www.nap.edu/openbook.php?record_id=12687&page=40) (April 2011).
- <sup>6</sup> Nilsen, K. (2009) 'Enhancing Access to Government Information: Economic Theory as it Applies to Statistics Canada,' in Paul Uhler (ed.) *The socioeconomic effects of public sector information on digital networks: Towards a better understanding of different access and reuse policies*, National Academies Press, Washington DC. Available [http://www.nap.edu/openbook.php?record\\_id=12687&page=40](http://www.nap.edu/openbook.php?record_id=12687&page=40) (April 2011).
- <sup>7</sup> Quiggin, J. (2010) *The value of Public Sector Information for Cultural Institutions*, Report to the Gov 2.0 Taskforce. Available <http://gov2.net.au/projects/project-6/> (May 2011).
- <sup>8</sup> Poplin, A. (2010) Methodology for Measuring the Demand Geo-information Transaction Costs: Based on Experiments in Berlin, Vienna and Zurich, *International Journal of Spatial Data Infrastructures Research*, 2010, Vol.5, 168-193. Available <http://ijmdir.jrc.ec.europa.eu/index.php/ijmdir/issue/view/20> (April 2011).
- <sup>9</sup> A summary to early 2008 can be found in Uhler, P.F., Sharif, R.M. and Merz, T. (2009) 'Measuring the Social and Economic Costs and Benefits of Public Sector Information Online: A Review of the Literature and Future Direction,' in Paul Uhler (ed.) *The socioeconomic effects of public sector information on digital networks: Towards a better understanding of different access and reuse policies*, National Academies Press, Washington DC. Available [http://www.nap.edu/openbook.php?record\\_id=12687&page=40](http://www.nap.edu/openbook.php?record_id=12687&page=40) (April 2011).
- <sup>10</sup> PIRA (2000) *Commercial exploitation of Europe's public sector information*, European Commission, Brussels. Available



- [http://www.epsiplus.net/psi\\_library/reports/commercial\\_exploitation\\_of\\_europe\\_s\\_public\\_sector\\_information\\_pira\\_study](http://www.epsiplus.net/psi_library/reports/commercial_exploitation_of_europe_s_public_sector_information_pira_study) (April 2011).
- 11 Weiss, P. (2002) *Borders in Cyberspace: Conflicting Public Sector Information Policies and their Economic Impacts*, US Department of Commerce, Washington DC. Available [http://www.epsiplus.net/psi\\_library/reports/borders\\_in\\_cyberspace](http://www.epsiplus.net/psi_library/reports/borders_in_cyberspace) (April 2011).
  - 12 Dekkers, M., Polman, F., te Velde, R. and de Vries, M. (2006) *MEPSIR: Measuring European Public Sector Information Resources*, European Commission, Brussels. Available [http://www.epsiplus.net/psi\\_library/reports/mepsir\\_measuring\\_european\\_public\\_sector\\_resources\\_report](http://www.epsiplus.net/psi_library/reports/mepsir_measuring_european_public_sector_resources_report) (April 2011).
  - 13 Te Velde, R. (2009) 'Public Sector Information: Why Bother?', in Paul Uhler (ed.) *The socioeconomic effects of public sector information on digital networks: Towards a better understanding of different access and reuse policies*, National Academies Press, Washington DC. Available [http://www.nap.edu/openbook.php?record\\_id=12687&page=40](http://www.nap.edu/openbook.php?record_id=12687&page=40) (April 2011).
  - 14 DotEcon (2006) *The commercial use of public information (CUPI)*, Report oft861, Office of Fair Trading, London. Available <http://www.oft.gov.uk/OFTwork/publications/publication-categories/reports/consumer-protection/oft861> (April 2011).
  - 15 Graves, A. (2009) 'The price of everything but the value of nothing,' in Paul Uhler (ed.) *The socioeconomic effects of public sector information on digital networks: Towards a better understanding of different access and reuse policies*, National Academies Press, Washington DC. Available [http://www.nap.edu/openbook.php?record\\_id=12687&page=40](http://www.nap.edu/openbook.php?record_id=12687&page=40) (April 2011).
  - 16 ACIL Tasman (2008) *The Value of Spatial Information*, Spatial Information Systems Limited. Available [www.crcsi.com.au/uploads/publications/PUBLICATION\\_324.pdf](http://www.crcsi.com.au/uploads/publications/PUBLICATION_324.pdf) (April 2011).
  - 17 OECD (2006) *Digital Broadband Content: Public Sector Information and Content*, OECD, Paris. Available [http://www.oecd.org/document/62/0,3746,en\\_2649\\_34223\\_32160190\\_1\\_1\\_1\\_1.00.html](http://www.oecd.org/document/62/0,3746,en_2649_34223_32160190_1_1_1_1.00.html) (April 2011).
  - 18 <http://www.oecd.org/dataoecd/41/52/44384673.pdf>.
  - 19 Government 2.0 Taskforce (2010) *Engage: Getting on with Government 2.0*, Department of Finance and Deregulation, Canberra. Available <http://www.finance.gov.au/publications/gov20taskforcereport/index.html> (April 2011).
  - 20 Australian Government (2010) *Declaration of Open Government*, 16 July 2010, Department of Finance and Deregulation, Canberra. Available <http://agimo.govspace.gov.au/2010/07/16/declaration-of-open-government/> (February 2011).
  - 21 <http://roarmap.eprints.org/>.
  - 22 For a full statement of principles, see OAIC (2011) *Principles on open public sector information*, OAIC, Canberra. Available [http://www.oaic.gov.au/infopolicy-portal/reports\\_infopolicy.html](http://www.oaic.gov.au/infopolicy-portal/reports_infopolicy.html) (June 2011).
  - 23 <http://eaves.ca/2009/09/30/three-law-of-open-government-data/>.
  - 24 Beagrie, N., Lavoie, B. and Woollard, M. (2010) *Keeping Research Data Safe 2*, IISC, London and Bristol. Available <http://beagrie.com/klds-i2s2.php> (May 2011).
  - 25 DotEcon (2006) *The commercial use of public information (CUPI)*, Report oft861, Office of Fair Trading, London, Annexe G, p34. Available <http://www.oft.gov.uk/OFTwork/publications/publication-categories/reports/consumer-protection/oft861> (April 2011).
  - 26 Houghton, J.W. (2009) 'Exploring the impacts of enhanced access to publicly funded research,' in Paul Uhler (ed.) *The socioeconomic effects of public sector information on digital networks: Towards a better understanding of different access and reuse policies*, National Academies Press, Washington DC. Available [http://www.nap.edu/openbook.php?record\\_id=12687&page=40](http://www.nap.edu/openbook.php?record_id=12687&page=40) (April 2011).

- 27 For details of the method see Houghton, J.W. and Sheehan, P. (2009) 'Estimating the potential impacts of open access to research findings,' *Economic Analysis and Policy* 39(1). Available [http://www.eap-journal.com/vol\\_39\\_iss\\_1.php](http://www.eap-journal.com/vol_39_iss_1.php) (April 2011); and Houghton, J.W., Rasmussen, B., Sheehan, P.J., Oppenheim, C., Morris, A., Creaser, C., Greenwood, H., Summers, M. and Gourlay, A. (2009) *Economic Implications of Alternative Scholarly Publishing Models: Exploring the Costs and Benefits*, Report to The Joint Information Systems Committee (JISC) by Victoria University & Loughborough University. Available <http://www.cfses.com/EI-ASPM/> (April 2011).
- 28 OECD (2006) *Digital Broadband Content: Public Sector Information and Content*, OECD, Paris. Available [http://www.oecd.org/document/62/0,3746,en\\_2649\\_34223\\_32160190\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/62/0,3746,en_2649_34223_32160190_1_1_1_1,00.html) (April 2011).
- 29 Productivity Commission (2001) *Cost Recovery by Government Agencies*, Report No 15, Canberra. Available <http://www.pc.gov.au/projects/inquiry/costrecovery/docs/finalreport> (May 2011).
- 30 Commonwealth of Australia (2005) *Australian Government Cost Recovery Guidelines*, Department of Finance and Administration, Canberra. Available [http://www.finance.gov.au/publications/finance-circulars/2005/09.html#FMG\\_4](http://www.finance.gov.au/publications/finance-circulars/2005/09.html#FMG_4) (April 2011).
- 31 Australian Government Cost Recovery Guidelines, Finance Circular No. 2005/09. Available [http://www.finance.gov.au/publications/finance-circulars/2005/09.html#FMG\\_4](http://www.finance.gov.au/publications/finance-circulars/2005/09.html#FMG_4) (May 2011).
- 32 Government 2.0 Taskforce (2010) *Engage: Getting on with Government 2.0*, Department of Finance and Deregulation, Canberra. Available <http://www.finance.gov.au/publications/gov20taskforcereport/index.html> (April 2011).
- 33 Tam, S-M. (2009) *Informing the Nation: Open Access to Statistical Information in Australia*, UNECE, Warsaw, May 2009. Available <http://www.unece.org/stats/documents/ece/ces/ge.45/2009/wp.11.e.pdf> (April 2011).
- 34 Tam, S-M. (2009) *Informing the Nation: Open Access to Statistical Information in Australia*, UNECE, Warsaw, May 2009. Available <http://www.unece.org/stats/documents/ece/ces/ge.45/2009/wp.11.e.pdf> (April 2011).
- 35 Treasury (2005) Press Release: *Improving Australia's Official Statistics*, 10 May 2005. Available <http://ministers.treasury.gov.au/DisplayDocs.aspx?doc=pressreleases/2005/051.htm&pageID=003&min=phc&Year=2005&DocType=0> (July 2011).
- 36 Treasury (2005) *Budget Paper No. 2, Budget Measures 2005-06*. Available <http://www.budget.gov.au/2005-06/pbs/html/abs-02.htm> (July 2011).
- 37 ABS (2010) *Extract from proposal to apply CC licensing to customised data output (January 2010)*, Canberra.
- 38 ABS (2007) *Impact of Free Stats on the Web on ABS Website Usage*, Canberra.; Fitzgerald , B. (2007) *An Australian Perspective on Open Access to and the Pricing of Public Sector Information*. Available <http://cci.edu.au/publications/digital-liberty-2007-publications> (May 2011).; Pwc (2010) *Economic Assessment of Spatial Data Pricing and Access*, Stage 1 Report, ANZLIC, Canberra. Available <http://www.anzlic.org.au/get/news/2744-1.pdf> (May 2011).
- 39 ABS (2007) *Impact of Free Stats on the Web on ABS Website Usage*, Canberra.
- 40 Unfortunately, ABS were unable to source the data required to reproduce this chart for ligibility.
- 41 ABS (2010) *Annual Report 2009-10*, Canberra, p109.
- 42 DotEcon (2006) *The commercial use of public information (CUPI)*, Report oft861, Office of Fair Trading, London, Annexe G, p34. Available <http://www.oft.gov.uk/OFTwork/publications/publication-categories/reports/consumer-protection/oft861> (April 2011).
- 43 DotEcon (2006) *The commercial use of public information (CUPI)*, Report oft861, Office of Fair Trading, London, Annexe G, p22. Available <http://www.oft.gov.uk/OFTwork/publications/publication-categories/reports/consumer-protection/oft861> (April 2011).



- 44 Houghton, J.W. (2009) 'Exploring the impacts of enhanced access to publicly funded research,' in Paul Uhler (ed.) *The socioeconomic effects of public sector information on digital networks: Towards a better understanding of different access and reuse policies*, National Academies Press, Washington DC. Available [http://www.nap.edu/openbook.php?record\\_id=12687&page=40](http://www.nap.edu/openbook.php?record_id=12687&page=40) (April 2011).
- 45 Houghton, J.W. and Sheehan, P. (2009) 'Estimating the potential impacts of open access to research findings,' *Economic Analysis and Policy* 39(1). Available [http://www.eap-journal.com/vol\\_39\\_iss\\_1.php](http://www.eap-journal.com/vol_39_iss_1.php) (April 2011); and Houghton, J.W., Rasmussen, B., Sheehan, P.J., Oppenheim, C., Morris, A., Creaser, C., Greenwood, H., Summers, M. and Gourlay, A. (2009) *Economic Implications of Alternative Scholarly Publishing Models: Exploring the Costs and Benefits*, Report to The Joint Information Systems Committee (JISC) by Victoria University & Loughborough University. Available <http://www.cfses.com/El-ASPM/> (April 2011).
- 46 Houghton, J.W. and Sheehan, P. (2009) 'Estimating the potential impacts of open access to research findings,' *Economic Analysis and Policy* 39(1). Available [http://www.eap-journal.com/vol\\_39\\_iss\\_1.php](http://www.eap-journal.com/vol_39_iss_1.php) (April 2011).
- 47 Useful reviews of this literature include Griliches, Z. (1995) 'R&D and productivity: Econometric Results and Measurement Issues,' In Stoneman, P. (Ed.) (1995) *Handbook of The Economics of Innovation and Technological Change*. Blackwell, Oxford, pp52–89.; Salter, A.J. and Martin, B.R. (2001) 'The economic benefits of publicly funded basic research: a critical review,' *Research Policy* 30(3), pp509-532.; Scott, A., Steyn, G., Geuna, A., Brusoni, S. and Steinmueller, E. (2002) *The Economic Returns to Basic Research and the Benefits of University-Industry Relationships*. Report to the Office of Science and Technology, London.; Shanks, S. and Zheng, S. (2006) *Econometric modeling of R&D and Australia's productivity*, Staff Working Paper, Productivity Commission, Canberra.; Martin, B.R. and Tang, P. (2007) *The benefits of publicly funded research*, SWEPS Paper No. 161, Science Policy Research Unit, University of Sussex, Brighton.; Sveikauskas, L. (2007) *R&D and Productivity Growth: A Review of the Literature*, US Bureau of Labor Statistics Working Paper 408, BLS, Washington, DC.; Hall, B.H., Mairesse, J. and Mohnen, P. (2009) *Measuring the returns to R&D*, NBER Working Paper 15622, NBER, Cambridge MA.; Hall, B.H., Mairesse, J. and Mohnen, P. (2010) *Measuring the returns to R&D*, in eds. Hall, B.H. and Rosenberg, N. (2010) *Handbook of the Economics of Innovation*, North Holland.
- 48 Department of Finance and Deregulation (2011) *Strategic Review of Geoscience Australia*, Canberra, pp97-98.
- 49 Department of Finance and Deregulation (2011) *Strategic Review of Geoscience Australia*, Canberra, pp97-98.
- 50 Fitzgerald, A.M. (2010) *State of Play: PSI Reuse in Australia*, European PSI Platform, Topic Report No. 13. Available [http://www.epsiplus.net/topic\\_reports](http://www.epsiplus.net/topic_reports) (April 2011).
- 51 At the end of March 1999, there were 14 nodes within the ASDD with a total of 19,774 dataset descriptions available for searching.
- 52 See <http://asdd.ga.gov.au/asdd/tech/quarterly.html>.
- 53 [http://water.gov.au/IntroductiontoAWR2005/index.aspx?Menu=Level1\\_2](http://water.gov.au/IntroductiontoAWR2005/index.aspx?Menu=Level1_2).
- 54 Water Information included any raw data and metadata, or any value added information product, that relates to: (a) the availability, distribution, quantity, quality, use, trading or cost of water; or (b) water access rights, water delivery rights or irrigation rights; and includes contextual information relating to water (such as land use information, geological information and ecological information).
- 55 <http://www.bom.gov.au/water/nwa/index.shtml>.
- 56 BOM (2010) *Improving Water Information: Progress Report*, BOM, Canberra. Available <http://reg.bom.gov.au/water/about/publications/index.shtml> (April 2011).
- 57 Watson, A. (1997) *Pricing Principles for CMSA Hydrologic Data and Information*, Melbourne.
- 58 Consultation with Stuart Minchin, CSIRO.

- 59 <http://www.bom.gov.au/water>.
- 60 BOM (2010) *Improving Water Information: Progress Report*, BOM, Canberra. Available <http://reg.bom.gov.au/water/about/publications/index.shtml> (April 2011).
- 61 DotEcon (2006) *The commercial use of public information (CUPI)*, Report oft861, Office of Fair Trading, London, Annexe G. Available <http://www.oft.gov.uk/OFTwork/publications/publication-categories/reports/consumer-protection/oft861> (April 2011).
- 62 ACIL Tasman (2008) *The Value of Spatial Information*, Spatial Information Systems Limited. Available [www.crcsi.com.au/uploads/publications/PUBLICATION\\_324.pdf](http://www.crcsi.com.au/uploads/publications/PUBLICATION_324.pdf) (April 2011).
- 63 This summary draws on Department of Finance and Deregulation (2011) *Strategic Review of Geoscience Australia*, Canberra.
- 64 PIRA (2000) *Commercial exploitation of Europe's public sector information*, European Commission: Brussels. Available [http://www.epsiplus.net/psi\\_library/reports/commercial\\_exploitation\\_of\\_europe\\_s\\_public\\_sector\\_information\\_pira\\_study](http://www.epsiplus.net/psi_library/reports/commercial_exploitation_of_europe_s_public_sector_information_pira_study) (April 2011).
- 65 Dekkers, M., Polman, F., te Velde, R. and de Vries, M. (2006) *MEPSIR: Measuring European Public Sector Information Resources*, European Commission, Brussels. Available [http://www.epsiplus.net/psi\\_library/reports/mepsir\\_measuring\\_european\\_public\\_sector\\_resources\\_report](http://www.epsiplus.net/psi_library/reports/mepsir_measuring_european_public_sector_resources_report) (April 2011).
- 66 Te Velde, R. (2009) 'Public Sector Information: Why Bother?', in Paul Uhler (ed.) *The socioeconomic effects of public sector information on digital networks: Towards a better understanding of different access and reuse policies*, National Academies Press, Washington DC. Available [http://www.nap.edu/openbook.php?record\\_id=12687&page=40](http://www.nap.edu/openbook.php?record_id=12687&page=40) (April 2011).
- 67 Graves, A. (2009) 'The price of everything but the value of nothing,' in Paul Uhler (ed.) *The socioeconomic effects of public sector information on digital networks: Towards a better understanding of different access and reuse policies*, National Academies Press, Washington DC. Available [http://www.nap.edu/openbook.php?record\\_id=12687&page=40](http://www.nap.edu/openbook.php?record_id=12687&page=40) (April 2011).
- 68 <http://www.oecd.org/dataoecd/9/61/38500813.pdf>.
- 69 <http://www.oecd.org/dataoecd/41/52/44384673.pdf>.
- 70 Beagrie, N., Chruszcz, J. and Lavoie, B. (2008) *Keeping Research Data Safe*, JISC, London and Bristol. Available <http://www.beagrie.com/klds.php> (April 2011).
- 71 Beagrie, N. (2009) *Draft Guide to Cost/Benefit Analysis for Research Data Services*, Charles Beagrie, Salisbury. Available [http://www.beagrie.com/DMIcost&benefit\\_programmeguidev1.pdf](http://www.beagrie.com/DMIcost&benefit_programmeguidev1.pdf) (May 2011).
- 72 Beagrie, N., Lavoie, B. and Wollard, M. (2010) *Keeping Research Data Safe 2*, JISC, London and Bristol. Available <http://www.beagrie.com/jisc.php> (April 2011).
- 73 Fry, J., Lockyer, S., Oppenheim, C., Houghton, J.W. and Rasmussen, B. (2008) *Identifying benefits arising from the curation and open sharing of research data produced by UK Higher Education and research institutes*, JISC, London and Bristol. Available <http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/jiscdataproposal-public.pdf> (March 2011).
- 74 Summary derived from Beagrie, N. (2009) *Draft Guide to Cost/Benefit Analysis for Research Data Services*, Charles Beagrie, Salisbury. Available [http://www.beagrie.com/DMIcost&benefit\\_programmeguidev1.pdf](http://www.beagrie.com/DMIcost&benefit_programmeguidev1.pdf) (May 2011).
- 75 <http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/jiscdataproposal-public.pdf>.