Impacts of increased access and new modes of consumption

SEB14 Session 1

John Houghton
Victorian Institute for Strategic Economic Studies
John.Houghton@vu.edu.au

Contents

Abstract

Examining the open data policy context we suggest that, while countries confront their own particular issues and difficulties in going from policy to practice and progress is uneven, there is a trend towards more open data (i.e. data made available at zero or marginal cost, with unrestrictive licensing, and machine readable in standardised data formats). There is also a shift towards being more inclusive, especially in the increasing expansion into static content and the GLAM sector (i.e. Galleries, Libraries, Archives, and Museums). We also note that there may be a wider open data opportunity, based around establishing a regulatory and policy environment conducive to enabling the maximization of opportunities to open private data – both business data and individual data.

Exploring the impacts of increased access, we note that the evidence suggests that providing open access to data leads to more use and creates greater value than do priced and/or restrictive access regimes. The principal dimensions of impact arising from increased access include:

- *Transaction cost savings* for both data providers and users;
- *Efficiency impacts* that can be quantified as time-cost savings;
- *Additional uses* that would not otherwise have been possible, thereby increasing the return on investment in the data collection/creation;
- Potentially *significant wider impacts* that are more difficult, perhaps impossible, to measure; and
- *New applications and combinations* of data, with unforeseen value and impacts emerging over time.

Looking at new modes of consumption, we discuss the impacts of online delivery of digital content, the increasing length and complexity of value chains, and the expanding 'shared base'. Noting that data are typically an intermediate good, with many and multiple possible uses and combinations of data feeding into many possible production processes, planned and serendipitous, we suggest that market signals are becoming more varied, fragmented, mediated, etc., and much less clear. Moreover, we suggest that the theoretical underpinnings of the private versus public trade-off shifts, with a larger role for government in the digital economy.

Exploring temporal analysis and changes in value and impact over time, we note many challenges facing those seeking to estimate the value of data and suggest that Thompson's 'Rubbish Theory' may resonate in this era of potentially unlimited use and re-use of data. Turning to the skills and capacity of data analysts and users, we identify some of the issues and opportunities arising, and the possibility of learning from previous examples of such challenges throughout the 'ICT Revolution'.

1. Introduction

This paper begins with a brief review of the policy context, looking at the scope of open data policies, recent policy developments and implementation progress, in order to identify major trends in open data policy and practice. Its focus then turns to the impacts of increased access and new modes of consumption.

Section 3 presents an exploration of the impacts of increased access to data, briefly touching on the theory behind open data, before examining the evidence and identifying the main dimensions of impact. Section 4 explores the impacts of new modes of consumption, focusing on the impacts of online delivery of digital content, the increasing length and complexity of value chains, the expanding 'shared base' and increasing role for government. Sections 5 and 6 explore temporal analysis and changes in value and impact over time, from both evidential and theoretical perspectives, and the skills and capacity of data analysts and users.

2. Policy context and developments

While the focus of the meeting is on geo-spatial information, it is worth considering the broader open data context. Insights and innovations often arise around the edges. Hence, this section explores the scope and development of open data policies, and briefly touches on implementation progress, before identifying the key and emerging trends.

2.1 The scope of open data policies

In discussion of open data, it is common to distinguish between types of data, both in terms of its sector of production and characteristics. The main focus of much of the earliest attention on open data was on government data or public sector information (PSI). A parallel focus has been on open access to research publications and data, particularly those arising from publicly-funded research. Nevertheless, the production of many types of data can span various sectors, with geo-spatial data being an example in which fundamental data, such mapping, cadastral and land use data, may be produced in the public sector, while many private sector entities may produce transport networks, points of interest locations, vehicle tracking, and other types of geo-spatial information. Similarly, research data may be produced in the public or private sectors.

There has also been an important distinction between dynamic and static data (Vickery 2014):

- *Dynamic* data refers to data that are continually generated. Much public sector information (PSI) is dynamic, being continually generated as a part of the functioning of government (e.g. national statistics, business registries, meteorological and geo-spatial data). These data are often readily usable in commercial applications with relatively little transformation of the raw data, as well as being the basis for extensive elaboration.
- *Static* data refers to data that are part of an established record or collection, which may be held by public sector agencies, but may not have been created by them. Static data collections are typically the province of the GLAM sector (i.e. galleries, libraries,

archives, and museums), although, the mapping is not perfect, nor is the distinction between static and dynamic always entirely clear.

Much research data is dynamic, in the sense of being continually produced as a part of the research process, but it can also take the form of one-off project data and form the basis of static collections. For example, site-based archaeological information may be the basis of a museum's collection.

2.2 Open data policy developments

There have been many initiatives at both national and international levels that seek to make more PSI, government and publicly-funded research data openly available. The basic rationale being that information can be a powerful tool in economic growth and competitive advantage.

Facilitating better access to PSI received widespread attention in Europe following the 2003 EU Directive on the re-use of public sector information. In 2008, the OECD Recommendation on Public Sector Information provided policy guidelines designed to improve access and increase use of PSI through greater transparency, enhanced competition and more competitive pricing (OECD 2008). The OECD 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 observed best practice and became the basis for further developments in PSI access regimes in OECD countries and elsewhere around the world.

For example, the G8 countries adopted an Open Data Charter in 2013, declaring that:

> *We, the G8, agree that open data are an untapped resource with huge potential to encourage the building of stronger, more interconnected societies that better meet the needs of our citizens and allow innovation and prosperity to flourish.*
>
> *We therefore agree to follow a set of principles that will be the foundation for access to, and the release and re-use of, data made available by G8 governments. They are: Open Data by Default, Quality and Quantity, Useable by All, Releasing Data for Improved Governance, and Releasing Data for Innovation* (Cabinet Office 2013).

An even wider range of developed and developing countries has been signing up to the Open Government Partnership (OGP). Launched in 2011, OGP seeks to provide an international platform for those committed to making their governments more open, accountable, and

responsive to citizens. Since 2011, OGP membership has grown from 8 to 64 participating countries.[1]

Similar policy initiatives have been adopted for research data, focusing on the data produced by publicly-funded research. In 2004, for example, governments of the [then] 30 OECD countries, together with China, Israel, Russia and South Africa, adopted a Declaration on Access to Research Data from Public Funding, which sought to maximise the return on public investment in research through open science data (OECD 2007).

Many national and charitable research funding agencies are also implementing policies that seek to make the publications and data arising from their funded research openly available (e.g. Research Councils UK, Australian Research Council, Wellcome Trust), either through the deposit of research papers in openly accessible subject and/or institutional repositories (e.g. the National Institutes of Health's PubMed Central), or through publication in open access journals and with open access publishers (e.g. the Public Library of Science (PLOS)). There is also an increasing range of data repositories and data journals emerging to facilitate the curation and sharing of data.

2.3 Progress towards open data

At the national level, there have been many initiatives. As Lateral Economics (2014) noted, the UK's 2007 Power of Information Report was a landmark, as was the election of Barack Obama in the US who signed a Memorandum on Transparency and Open Government on his first day in office. Australia was a fast follower, and widely praised at the time of its 2009 Government 2.0 Taskforce. The intervening years have seen progress continue, albeit unevenly, with countries bringing their own focus and leading in particular areas. Examples include Denmark's comprehensive approach to basic government data, and the UK's 'midata' initiative[2] attempting to foster greater understanding, trust and involvement through individual control.

Table 1 shows the top 20 countries on the Open Data Barometer ranking in 2013. As with other such rankings, the United States, United Kingdom, Scandinavian countries, Australia, New Zealand and Canada tend to top the list, though all have their strengths and weaknesses. For example, Ordnance Survey Great Britain only provides its medium and small scales geo-spatial data free at the point of use, while covering its operating costs through the sale of large scale data.[3]

Open government data initiatives have often been underpinned by the use of centralised web sites to distribute government data and information, beginning with the United States (data.gov, May 2009), the United Kingdom (data.gov.uk, September 2009), New Zealand (data.govt.nz, November 2009), Norway (data.norge.no, April 2010), Australia (data.gov.au, March 2011), Canada (data.gc.ca, March 2011), and subsequently followed by many other countries (Vickery 2014, p7). There are also national foci and centralised web sites for research data curation and

1 http://www.opengovpartnership.org/

2 http://www.midatalab.org.uk/midata-explained/

3 http://www.ordnancesurvey.co.uk/docs/annual-reports/ordnance-survey-annual-report-and-accounts-2013-14.pdf

sharing (e.g. the UK Data Archive and the Australian National Data Service's Research Data Australia).[4]

Table 1 Top 20 Countries by their Open Data Barometer Ranking, 2013

Country	*Readiness*	*Implementation*	*Impact*	*ODB-Scaled*	*Rank*
United Kingdom	0.50	3.09	0.71	100	1
United States	0.46	2.49	0.94	93	2
Sweden	0.45	2.32	0.62	86	3
New Zealand	0.33	1.52	0.82	74	4
Denmark	0.34	1.73	0.43	72	5
Norway	0.42	1.77	0.32	72	5
Australia	0.38	1.48	0.38	68	7
Canada	0.30	1.45	0.38	66	8
Germany	0.26	1.41	0.41	65	9
France	0.30	1.47	0.24	64	10
Netherlands	0.37	1.59	0.03	64	10
Korea (Rep. of)	0.28	1.04	0.07	54	12
Iceland	0.15	0.95	0.09	51	13
Estonia	0.24	0.79	0.06	49	14
Finland	0.42	0.41	0.26	49	14
Japan	0.28	0.68	0.11	49	14
Spain	0.19	0.79	0.03	48	17
Austria	0.20	0.32	0.35	46	18
Israel	0.14	0.62	0.08	46	18
Italy	0.03	0.48	0.31	45	20
Russia	0.06	0.37	0.35	45	20

Source: http://www.opendataresearch.org/content/2013/535/get-data-open-data-barometer-2013

International organizations have also launched open data portals. For example, the United Nations' portal (data.un.org) was launched in 2008, earlier than the US government's data.gov. The OECD's portal (OECDiLibrary) was launched in 2010, offering open access to publications and data from the OECD, International Energy Agency (IEA), Nuclear Energy Agency (NEA), Programme for International Student Assessment (PISA), and International Transport Forum (ITF). The World Bank's repository (data.worldbank.org) was launched in 2011 and has been praised for its (human and machine) usability. The European Commission's portal (open-data.europa.eu) was launched in 2012 hosting datasets from European organizations, including Eurostat (Science-Metrix 2013, p6).

2.4 Trends in policy and practice

Countries confront their own particular issues and difficulties in going from policy to practice. The difficulties of overcoming resistance from historical cost recovery schemes and privatised agencies (e.g. the UK's Trading Funds), and the many practical difficulties encountered in making data available (e.g. Denmark's Business Registry included free text fields that were not

4 http://www.data-archive.ac.uk/ and https://researchdata.ands.org.au/

machine readable). Nevertheless, *the trend is towards more open data* – data made available at zero or marginal cost, with unrestrictive licensing, and machine readable in standardised data formats.

There is also a shift towards being more inclusive, especially in the increasing expansion into static content and the GLAM sector. For example, the 2003 European Directive on the re-use of public sector information excluded information and content generated and held by cultural and educational institutions and public sector broadcasters; whereas the 2008 OECD Recommendation of the Council for enhanced access and more effective use of PSI includes all information and content generated by and/or held by public bodies (Vickery 2014, p6).

There is also increasing awareness of a wider open data opportunity, as the increasing ease with which data can be stored, used and distributed increases the marginal benefit of all such activities. Hence, there are wider opportunities for government action. In a report for the Omidyar Network, Lateral Economics (2014, p79) outlined a number of these wider opportunities, based around:

- Direct open data policies focusing on making more government data openly available;
- Supporting and prosecuting policies focusing on opening publicly-funded research publications and data; and
- Establishing a regulatory and policy environment conducive to enabling the maximization of opportunities to open private data – both personal data and business data.

3. Impacts of increased access

To explore the impacts of increased access to data, this section begins with a very brief review of the theory before turning to focus on the evidence of impacts arising from increasingly open data, including PSI and government data, as well as publicly-funded research publications and data.

3.1 The theory behind open data

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 (Houghton 2011). While information can be made more or less excludable through intellectual property rights, 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 involvement in the supply of information (Nilsen 2007; 2009).

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 – as 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 and access is close to zero. Cost recovery through such mechanisms as user fees

is unlikely to be welfare enhancing (Nilsen 2007; 2009). For most PSI, government and publicly-funded research data, marginal cost pricing will be the preferable option (Pollock 2009, p40).

Some public sector agencies retain copyright in the information they produce to enable them to control access and the conditions of use, and to prevent uses that would result in the loss of control. However, if the price is set to the marginal cost of online dissemination (zero), there is no purpose to be served by restrictive licensing. Indeed, as the costs of disseminating and accessing information have declined, the transaction costs associated with controlling access to information have come to constitute a significant barrier to access in themselves (Quiggin 2010). Studies of these transaction costs suggest that they can be significant (Poplin [nee Krek] 2010). Hence the efficient economic solution for the dissemination of PSI and government data, and publicly-funded research is likely to be free *libre* and free *gratis* (i.e. making the data freely available online using unrestrictive licensing, such as Creative Commons).

3.2 Evidence on the impacts of increased access

The evidence suggests that providing open access to data (i.e. access that is free or at marginal cost, where unrestrictive licenses apply, and it is provided in standardised machine readable formats) leads to more use and creates greater value than do priced and restrictive access regimes. This can be seen across data types and sectors.

3.2.1 Public Sector Information

There have been many studies showing that the value of PSI exceeds the costs, and how making PSI more openly available can significantly increase the benefits without greatly increasing costs.

For example, using a top-down approach, PIRA (2000) estimated the investment in, and economic value 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 – representing a seven-fold return on investment. By comparison, they put PSI investment in the United States at EUR 19 billion per annum and its economic value at EUR 750 billion – a 39-fold return on investment. They attributed the difference to the US having a more open access regime than EU countries, and suggested that the EU could reach US levels if they adopted open data regimes. Moreover, they estimated that EU countries would only need to double the value of PSI for governments to recoup lost revenues from PSI sales in increased tax receipts (PIRA 2000; Weiss 2002).

In a report to the UK Office of Fair Trading (OFT), DotEcon (2006) developed a bottom-up approach to estimating the 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. They also looked at the cost of major barriers to use ('detriment'), including unduly high prices, distortion of downstream competition, and failure to exploit PSI. 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.

Reviewing the literature on the size and growth of the European market for PSI, Vickery (2010) noted that the EU27 countries direct PSI-related market was of the order of EUR 28 billion in 2008. Noting that all studies show relatively rapid growth in PSI-related markets, and assuming annual growth of 7%, he suggested that the direct PSI-related market would be around EUR 32 billion in 2010. PSI-related information can be used in a wide range of direct and indirect applications across the economy, and aggregate direct and indirect economic impacts from PSI applications and use across the EU27 economies were estimated to be of the order of EUR 140 billion annually. These estimates were based on business as usual, but it was estimated that overall economic gains from opening up PSI and providing easy access for free or at the marginal cost of distribution, could be up to EUR 40 billion for the EU27, and aggregate direct and indirect economic benefits could have been of the order of EUR 200 billion in 2008 – more than 40% higher than in the 'business as usual' scenario.

Vickery (2010) went on to note that sectoral estimates of possible gains from the removal of current barriers to access and improving the underlying infrastructure include the geo-spatial sector, where benefits could be increased by some 10% to 40% by improving access, data standards, and building skills and knowledge. He suggested that better policies in the area of geo-spatial applications in local government could help productivity gains almost double over the next five years. In terms of gains in the efficiency of existing operations, improving accessibility of information necessary for obligatory environmental impact assessments could potentially reduce EU27 environmental compliance costs by 20% or around EUR 2 billion per year, open access to R&D results could result in recurring annual gains of around EUR 6 billion per year, and if European citizens each saved as little as two hours per year by more rapid and comprehensive access to PSI, it would be worth at least EUR 1.4 billion per year.

There are many such studies, reviews and summaries, in which estimates of the impacts of a future characterized by more open access are proffered (ex-ante). While fewer in number, there have also been studies exploring the impact of more open access after the event (ex-post). For example, in a study for the Australian National Data Service, Houghton (2011) explored the costs and benefits of free access to government data and standardised licensing and formats, using the cases of national statistics, fundamental geo-spatial data, and hydrological data. This is one of the few studies that brings together the direct agency and user impacts of open and transparent access with the wider economic impacts. It is also one of the few studies to compare the situation before and after the introduction of free and open access and standard licensing (i.e. to explore the real world both before and after the open data regimes were introduced).

Houghton (2011) found 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 AUD 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 AUD 5 million per annum. Moreover, there were significant agency savings of around AUD 1 million per annum (e.g. in handling subscriptions and sales, financial transaction fees, and user enquiries and

support), and the agency noted the managerial and mission benefits of freeing staff from such activities and being able to redeploy them to focus on the agency's core business.

Estimates of the impacts of making ABS publications and data free online on consumer welfare ranged from AUD 4 million to AUD 5 million per annum at 2005-06 prices and levels of activity. The wider impacts arising from additional use bring substantial additional returns, with estimates suggesting overall costs associated with free access to ABS publications and data online and unrestrictive standard licensing of around AUD 4.6 million per annum and measurable annualised return on investment benefits of perhaps AUD 25 million – more than five times the costs (Houghton 2011).

A crucial element in this latter estimate lay in identifying the change in use of the data between the toll access and open access regimes, using website statistics as a proxy indicator of use. 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 contemporaneous trends and influences, such as the impacts of:

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

In practice, the evidence regarding increasing use proved to be difficult to establish with certainty.

3.2.2 Science publishing

Recent years have seen the emergence of open access publishing, especially for publications arising from publicly-funded research. By some estimates, up to 48% of research articles published during 2008 were openly and freely available by the end of 2012 (Science-Metrix 2013).

Numerous studies of the possible advantages have sought to explore the impacts of open access on use, as indicated by citations and downloads. The supposed increase in citations has long been used as an argument to encourage research authors, and their funders, to adopt open access. Many, indeed the majority of studies have found a (sometimes significant) 'citation advantage'. For example, Swan (2010) reviewed 31 studies and found that 27 (87%) reported a citation advantage, and Wagner (2010) reviewed 46 studies and found that 39 (85%) did so. But citations are not the full story.

Citation is an indicator of use by other authors, who cite the works in their work. Many authors, such as those in universities and research centres, may have enjoyed reasonable access under the toll access regime. Consequently, it is important to measure additional use by those who do not themselves author new works and may have confronted more limited access (e.g. private sector researchers and research users/appliers). And there does seem to be widespread acceptance that

there is a 'download advantage', suggesting that use does increase even if citations do not. For example, in one of the relatively few studies reporting no citation advantage, Davis et al. (2008) noted an 89% increase in full-text downloads. More recently, RIN (2014, p6) analysed citation and download data for *Nature Communications*, concluding that: "the open access articles received significantly more usage than those published under a subscription model."

This evidence fits with evidence of the need for, and use of, academic research of the sort typically published in journal articles. For example, Ware (2009) looked at access by UK small and medium-sized enterprises (SMEs) to professional and academic literature, and found that SMEs rated original research articles and review papers in journals as their most important sources of information. Similarly, in a study of access to research and technical information among technology-based SMEs in Denmark, Houghton, Swan and Brown (2011) found that 68% of respondents reported reading or consulting research articles on a monthly or more regular basis, 45% on a weekly or more regular basis and 10% on a daily basis.

These studies also suggest unmet demand and access difficulties. For example, Ware (2009) found that 73% of SME respondents and 53% of large firm respondents reported having difficulties accessing research articles. Similarly, Houghton, Swan and Brown (2011) found that 38% of Danish SME respondents said they always or frequently had difficulty getting the research articles they needed, and a further 41% said they sometimes had difficulties. Therefore, it seems highly plausible that more open access leads to greater use.

Estimates of the possible benefits of open access publishing focus on activity-cost savings, efficiency impacts and possible increases in return on investment resulting from additional access and use. For example, in a series of studies comparing the cost of publishing research articles under alternative models and exploring the wider benefits of more open access to research (Houghton et al. 2009; Houghton 2009a, 2009b, 2009c; Houghton, de Jonge and van Oploo 2009; Houghton, Rasmussen and Sheehan 2010; Houghton et al. 2012; Houghton and Swan 2012, 2013), found that major savings include research time (i.e. efficiency), and research library acquisition and handling costs (i.e. activity and transaction costs). The other major impact was the additional return on investment in the research arising as a result of the increases in accessibility and the efficiency of the research process.

3.2.3 Research data

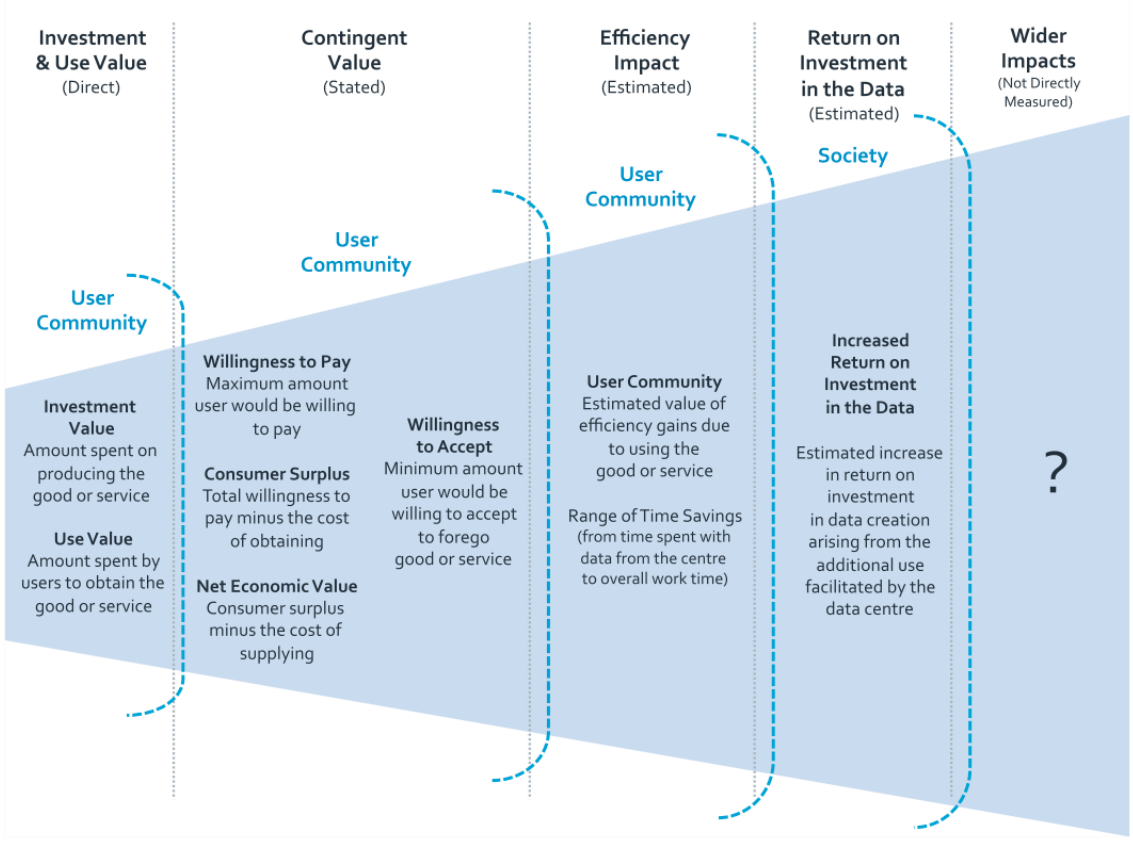
A similar pattern can be seen in studies of the impacts and benefits of curating and openly sharing research data, with many ex-ante studies suggesting and sometimes quantifying potential or expected benefits, but relatively few ex-post studies measuring the benefits that have been realised. Of course, as time passes, there is more scope for ex-post analysis.

Among qualitative studies, are the series of projects undertaken by UK-based Charles Beagrie Ltd under the general heading Keeping Research Data Safe (KRDS). The initial KRDS study 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 such preservation (Beagrie et al. 2008). A second phase project (KRDS2), further developed the activity-based cost model presented in the original study, presented detailed cost information for four organizations (i.e. the Archaeology Data Service, the National Digital

Archive of Datasets, the UK Data Archive, and the 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 (Beagrie et al. 2010). The study found that there can be significant benefits for current researchers in the short-term, as well as long-term benefits to future research.

A subsequent series of studies of UK-based research data centres combined the largely qualitative KRDS framework with a number of quantitative approaches to measure the value and impact of research data curation and sharing. The studies covered the Economic and Social Data Service (ESDS), the Archaeology Data Service (ADS), and the British Atmospheric Data Centre (BADC) (Beagrie et al. 2012; Beagrie and Houghton 2013a, 2013b, 2014). The economic methods used can be seen as estimating a range of values, moving from those focusing on minimum values toward methods that measure some of the wider impacts (Figure 1).

Figure 1 Methods for exploring the economic value and impacts of research data centres



Source: Beagrie, N. and Houghton, J.W. (2014) *The Value and Impact of Data Sharing and Curation: A Synthesis of Three Recent Studies of UK Research Data Centres*, Joint Information Systems Committee (Jisc), Bristol and London.

The methods include:

- Estimates of investment value (i.e. the operational expenditure of the data centres plus the time and other costs for depositors submitting data), and use value (i.e. the cost of the time spent by users accessing the data and services);
- Contingent valuation (i.e. willingness to pay and willingness accept), using stated preference techniques (DTLR 2002), to explore the amount that users would be willing to pay to access the data and services, or would be willing to accept in return for giving up their access, in a hypothetical market situation;
- Welfare approaches to estimating consumer surplus (i.e. willingness to pay minus use value) and the net economic value (i.e. consumer surplus minus operational budget) of the data and services provided by the data centres;[5]
- An activity-cost approach to exploring the estimated work-time saving (i.e. efficiency) impacts of the research data centres among their user communities; and
- A macro-economic approach, using a modified Solow-Swan model (Houghton and Sheehan 2009), to explore the increase in social returns[6] on investment in the original creation/collection of the data hosted, arising from the additional use of the data facilitated by the centres (i.e. the implied value of the data re-use by those who could neither have obtained the data elsewhere nor created/collected it themselves).

The economic analysis indicated that:

- A very significant increase in research, teaching and studying efficiency was reported by users as a result of their using the data centres. These estimated efficiency gains ranged from 2 and up to more than 20 times the costs – including operational, depositor and user costs.
- The value to users exceeded the investment made in data sharing and curation via the centres in all three cases. The indirect investment made in the data centres by depositors (and their funders), in terms of the time and other costs involved in preparing material for deposit, is substantial. Nevertheless, what users pay, in terms of their access time, and what they would be willing to pay for access, is 2.2 to 2.7 times greater than the value invested in the centres (i.e. in terms of operational costs plus depositor costs).
- And the estimated value of the increase in return on the original investment in the creation/collection of the data hosted, resulting from the additional use facilitated by the centres, ranged from twice and potentially up to 12 times the investment in the data centres (Beagrie and Houghton 2014).

In terms of value and impact, it is worth noting that the research data users came from all sectors and all fields. For example, around 20% of respondents to the ESDS user survey were

5 While value may be considered to include both consumer and producer surplus, the data and data centre services in these studies were free. Hence, there was no producer surplus to consider.

6 Social returns refer to the sum of private and public returns (i.e. both the returns that can be captured by the creator/user and those that spill over to others).

from government, non-profit and commercial sectors, as were around 40% of respondents to the BADC user survey, and close to 70% of respondents to the ADS users survey.

Nevertheless, these largely survey-based economic analyses confronted a number of difficulties in implementation. For example, there is the issue of weighting. The focus of the economic analysis is on the value of the data centre. But few, if any, users would use all of the data hosted and provided. Rather, they experience just a part of it. So the costs and value they report relate to the part they use, not to the centre as a whole. Therefore, it is necessary to weight the survey responses to reflect the overall pattern of use of the data and services. Such a weighting necessarily depends upon matching survey respondents' reported activities with weblog recordings of the overall activity on the data centre (Beagrie and Houghton 2014). Any discrepancies between activity perceptions and activities captured in weblogs are problematic.

3.2.4 New applications and combinations

There are also many new applications and combinations that bring value and increase the impact of data. Everyone has their favourite examples from 'hack days', and at the moment mine is Intec's 'When the heck am I?' app, which emerged at the 2014 Australian GovHack event.

Using historic images and information from multiple datasets and geo-location tags, the app lets users explore the heritage of a city using the Google Glass wearable device. When a user arrives near an historic point of interest, archived photographs recorded from or near the current location are displayed on Google Glass to present a visual history of the place. The app uses precise geo-spatial data to improve the relevance and usefulness of the vast catalogues of imprecise, historic photographs available – initially using Australian State and National Library archives, and focusing on Adelaide.[7]

New applications of this sort can bring renewed interest and value (e.g. to historical photographic collections), and the combinations (e.g. of historical photographic collections and geo-spatial data) can add value, making the value of the combination, in some senses, greater than the sum of the parts.

3.3 Summarising the impacts of increased access

These studies reveal the dimensions and nature of impacts arising from increased access to data. They show:

- *Transaction cost savings* for both data providers and users, to the extent that direct agency cost savings alone could be sufficient to cover the possible loss of revenue arising from providing open access. These activity time-cost savings can be quantified relatively easily;
- *Efficiency impacts*, primarily for data users, but also for suppliers, that can be quantified as time-cost savings. In some circumstances, these savings may be reinvested in the activity, thereby having further return on investment impacts (e.g. where research time

7 http://hackerspace.govhack.org/content/when-heck-am-i

saving results in more research being done, thereby increasing the return on investment in research);

- *Additional use* that would not otherwise have been possible, thereby increasing the return on investment in the data creation/collection (e.g. pure additional uses of data by those who could neither create/collect the data themselves nor obtain it from elsewhere);
- Potentially significant *wider impacts* that are more difficult, perhaps impossible, to measure (e.g. greater engagement and safety during outdoor activities as a result of combinations of location-specific information and geo-tracking); and
- *New applications and combinations*, with unforeseen value and impacts emerging over time.

Studies across information and data types, be they ex-ante or ex-post, suggest that the benefits of more open access (i.e. free or at marginal cost, with unrestrictive licensing, and standardised machine readable formats) substantially outweigh the costs.

4. New modes of consumption

In addition to important changes in modes of consumption as content is increasingly digital and delivery and consumption increasingly taking place online, there are also some important longer-term changes in the economic organization of both production and consumption.

4.1 Digital content and online delivery

A, perhaps the, major change in delivery and consumption is that it is increasingly featuring online digital content. The economic implications of this are well rehearsed, hence only briefly summarised here.

Digitalisation makes content-based goods even more like public goods, in terms of both being non-rivalrous in consumption (e.g. one person's accessing a website has no impact on others' access, whereas it is difficult for two people to read the same copy of a book simultaneously) and non-excludable (e.g. it is much easier and cheaper to create multiply perfect or near perfect copies of a digital music track than of a printed book). There are also inherent information asymmetries, as information and its expression are 'experience goods' (i.e. consumers cannot fully know their value until they have been consumed, so they are purchased without full knowledge of their value).

Traditional economic analysis examines situations where the relevant technology involves no economies of scope and constant or decreasing returns to scale. In such circumstances, the conventional wisdom 'set prices at marginal cost' is both economically viable and the likely outcome of competitive forces. However, digital content products typically involve technologies that exhibit increasing returns to scale (often in an extreme form where the marginal cost of production is close to zero), large fixed and sunk costs, and significant economies of scope. In such circumstances, setting prices equal to marginal cost is unlikely to recoup sufficient revenue

to cover the fixed costs, and the standard economic recommendation of 'price at marginal cost' may not be economically viable (Thorpe 2003).

The problem is that any price above marginal cost will exclude some from consuming who otherwise would, and is necessarily welfare reducing because the goods are being under-utilized. In static welfare terms:

> *Any price or restriction that does not reflect a seller's costs of getting the marginal unit to market causes a corresponding loss in value created – what economists refer to as a deadweight loss. A deadweight loss occurs when a seller, in order to extract more from some consumers, either prices to exclude other consumers or to reduce overall use of the product. When the seller genuinely incurs costs, prices can play a role in getting consumers to share in the efforts to economize on those costs. However, for information goods, digitization reduces the need for such economizing and, correspondingly, increases the size and potential for deadweight losses* (Gans 2012, p14).

Hence, digital content and online delivery bring challenges, as well as enormous opportunities.

4.2 The length and complexity of value chains

Nineteenth century economists referred to the 'roundaboutness' of production, noting the roundabout nature of emerging production processes and the increasing length and complexity of production chains. "A steel ladder, for example, will be produced and brought to market only if the demand supports the digging of iron ore, the smelting of steel, the machines that press that steel into ladder shape, the machines that make and help maintain those machines, etc."[8] Even then the ladder may be bought for use in their businesses by a window cleaner or painter and decorator, who in turn sell services that may be input to the building industry, and so on.

Over the last 20 years or so, there have been a number of trends contributing to ever greater 'roundaboutness', including globalisation, enabled in part by advances in information and communication technologies and the deregulation of transport and communications, and a shift from vertical to horizontal integration and from economies of scope to economies of scale in production (Houghton 2006). As the OECD (2005, pp16-17) put it:

> *"The globalisation of trade in goods and services is opening up new and increasingly large markets… One consequence of these changes is the fragmentation of production processes, with different stages of production carried out in different countries."*

Multinational firms have located different parts of their production processes around the world to take advantage of sometimes quite small differences in costs, resource availability, logistics and market access. What is distinctive about international production systems since the late 20th century is the intensity of integration on a global scale and the emphasis on the efficiency of the system as a whole (Kaplinsky 2000).

---

8 I do not refer here to the controversial and discredited theory of value that Böhm-Bawerk expounded through the notion of roundaboutness, but rather simply note the fact of roundaboutness (http://en.wikipedia.org/wiki/Roundaboutness).

Data are typically an intermediate good. Rarely, if ever, are data destined for final consumption. Moreover, as data are not exhausted in consumption, there never is an end user (or use) per se, but rather endless potential users and uses. There may also be serendipitous use (i.e. unplanned and unintended use). With many and multiple possible uses and combinations of data feeding into many possible production processes, planned and serendipitous, market signals are becoming much more varied, fragmented, mediated, etc., and much less clear.

4.3 An increasing shared base

There are a number of dimensions along which the 'shared base' has been expanding. In research, there has been a greater expansion of those sectors undertaking 'pre-competitive' research, and the emergence of 'open innovation' (Chesbrough 2003; Chesbrough et al. 2006). And, as noted, there has been an expansion of open access publication of research results and the open curation and sharing of research data. In electronics, there has been a push towards greater availability and use of shared IP cores, including OpenCores.[9] And, of course, there is ever greater use of open source software. In some key areas, such as internet server operations, open source software is close to ubiquitous.[10] Moreover, for an ever wider range of what are increasingly the most valuable economic activities, such as research, education, internet intermediary activities, etc., we are seeing an expansion of fair use / fair dealing, and exceptions to copyright (Houghton and Gruen 2012).

Perhaps the most important, certainly most visible, open platform for public goods is the internet, which is characterised by its multiple levels of openly accessible platforms – from communications and servers, to Google, Twitter, Facebook, and so on. These platforms are potentially excludable, but would not be anything like as valuable if excludability were enforced. As Gruen (2010) has said:

> *Google and Facebook could close their platforms and charge you for access to them. But… they would not be anywhere near as socially valuable if you were charged – because participants add value on social networking platforms. In fact, they add so much value that private profit seeking builders of such platforms leave them open to all. They generate such vast social value that way that if the builders can monetise just a small fraction of that value they can become rich beyond their wildest dreams.*[11]

4.4 An increasing role for government

Stiglitz et al. (2000) suggested 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, because information products have special characteristics that raise the likelihood that the market will fail to produce information at a socially optimal level.

---

9 http://opencores.org/

10 http://en.wikipedia.org/wiki/Open-source_software#Current_applications and http://opensource.com/resources/projects-and-applications

11 http://clubtroppo.com.au/2010/09/16/mr-gruen-goes-to-washington-again-2/

The US Department of Commerce (2014) recently argued that products that can be produced efficiently by the market share five features:

1. They are rivalrous, that is, only one economic agent can consume a given unit of such products;
2. They are easily excludable, such that the producer has property rights to compel a consumer to pay for the product;
3. They have characteristics and benefits that are open in the sense that they are easy to recognize or understand;
4. They generally have constant returns to scale or possibly decreasing returns to scale, at least above some level of output; and
5. The production and consumption of such products generally do not impose gains or losses on others (i.e. they generate no externalities).

In contrast, they suggested that information generally does not share these characteristics.[12] The consumption of information is non-rivalrous (i.e. one person's use of the information does not diminish another's use of it), and although information can be made to be excludable, it is often costly to do so. These two characteristics suggest that information has the characteristics of a 'public good'. Standard economic theory shows that markets generally provide public goods at less than the socially optimal level. As a result, government intervention could help the economy achieve this socially optimal level. Moreover, information often is opaque (i.e. its quality can be difficult to ascertain before acquiring it). This opacity can lead to lower than optimal production *and* consumption of information.

Information often exhibits increasing returns to scale, because of high fixed costs of production and negligible variable costs (e.g. low costs of dissemination). Moreover, information production can have economies of scope (i.e. the average cost of production declines as additional product types are produced), because some of the infrastructure of the production process for one type of information product may also be used to produce another information product. Such increasing returns to scale and scope give rise to natural monopoly conditions, with the result that the supply of information, if left to a monopoly market supplier, would be less than the socially optimal level.

The production and consumption of information can generate external benefits. For example, as more people, businesses, and governments use information to make better decisions, the information becomes more valuable. In addition, setting standards for data, such as definitions of key concepts, makes all data that employs such standards potentially more valuable. In this way, there are network externalities in the production and use of information. Given these externalities, private consumers and providers do not consider the full set of benefits when deciding how much information to exchange, and less than the socially optimal amount of information will be produced.

---

12 The following paragraphs draw heavily from US Department of Commerce (2014) *Fostering Innovation, Creating Jobs, Driving Better Decisions: The Value of Government Data*, Department of Commerce, Washington DC, pp20-21.

The characteristics of information (i.e. non-rivalry, costly excludability, opacity, economies of scale and scope, and the presence of positive externalities) imply that the market would be likely to produce less than the socially optimal level of information. As a result, there is a role for government intervention to help the economy achieve a more optimal level of information (US Department of Commerce 2014, pp20-21).

4.5 No such thing as a free lunch

Of course, there is no such thing as a free lunch. Indeed, it is often argued that there are dangers in 'free', and benefits in having market prices and signals, in terms of alignment of production with demand, quality, reliability, etc.

Arguing for PSMA Australia's[13] full cost recovery approach, Appleyard (2014, pp5-6) noted significant benefits for industry, saying:

- *The cost of developing national data is spread across the market, making the cost to an individual user or organisation radically lower than the investment required for them to produce the same thing.*
- *Market economics dictate that PSMA will only generate a return from its products if they align with market need and are available for an acceptable price. This guarantees that PSMA remains responsive and relevant to real market: being market-driven, PSMA remains market-focused.*
- *Also, competition and the threat of competition, drives efficiency in PSMA operations as well as continued investment in systems, infrastructure and processes.*
- *And importantly, PSMA's funding and sustainability remain independent of Government budgetary pressures and policy decisions, meaning it can provide its products and services to the market with continuity, stability and longevity – without the threat that government financial constraints may impact its funding levels, product development, product quality and continued viability.*

Appleyard (2014, p2) began his presentation saying: "we all agree that national spatial data is paramount… but who will pay for this data?" It is a good question, but I think it is better to start with an alternative question: Who won't pay? And why?

Willingness to pay is constrained by capacity to pay, and certain users may be excluded where capacity to pay is extremely limited. Perhaps more importantly, certain uses may be excluded where capacity to pay is limited and/or the value of the use is uncertain or unrealisable (e.g. spills over to others and cannot be fully captured), such as:

- Where the nature of use means that a number of pieces of information are required as inputs and willingness to pay for individual items is close to zero;

13 PSMA Australia Limited is a company owned by state, territory and Australian governments, established to coordinate the collection of fundamental national geo-spatial datasets and to facilitate access to this data (http://www.psma.com.au/).

- Where use is non-commercial, and the revenue generated from it is low or non-existent, limiting capacity to pay for the optimal level of use;
- Where use is transformative and the value is uncertain (e.g. innovative new products or services); and/or
- Where the use has positive externalities, such that the users do not capture the value of their use as it spills over to others.

Cost recovery cannot produce efficient outcomes in such circumstances, rather the benefits of these uses are simply lost while the incentive to produce remains unchanged. As the Danish Government (2012, p4) noted: the price of data may prevent both entrepreneurs and established businesses from testing the commercial opportunities associated with exploiting the data in new and creative ways.

In his analyses of public goods, Nicholas Gruen has pointed out that non-excludability creates a free-rider problem, but non-rivalrousness creates a free-rider opportunity, and in many cases the free-rider opportunity is much greater – as the benefits of not enforcing excludability greatly exceed those of enforcement. Using Google as an example, Gruen (2014) estimated that Google creates around USD 1 trillion in value a year, while getting by on a mere USD 60 billion in revenue, making Google around 94% a gift to the world. Google, and many other web platforms, have forgone defending themselves against the free-rider problem in order the grasp the free-rider opportunity.[14]

Who will pay? – producers of established products and services in the market sector, which do not exhibit public good characteristics and for which the data are a major input. Who won't pay? – innovative, integrative uses for which the outcome is uncertain (e.g. no established market) or is itself a public good (e.g. research and education, data/information), etc. As the Danish Government (2012, p16) put it:

> *The price of data, and rights to it, can be a barrier for new businesses who want to exploit data commercially... However, with open basic data, businesses can test new ideas at low risk, which leads to a great potential for innovation within the market; innovation which in turn generates growth and improved products for users.*

Moreover, the market is not the only, nor necessarily the best, mechanism for transmitting signals. In cases where willingness to pay is constrained, the market may not work. For example, the Danish Government (2012, p22) reports examples of government departments and utilities choosing cheaper/inferior data when they have to pay, saying: "As things are today, when the utilities buy data, they often choose the cheapest solutions, which means the data is not always up-to-date." And there are non-market alternatives. For example, the Australian Bureau of Statistics (ABS) makes extensive use of 'user groups' to get feedback on existing products and services and leads as to what is needed by their users – with 78 such groups active during 2013 (ABS 2013).

14 https://www.youtube.com/watch?v=7YmD1gg7txY

Box 1 A Framework for 'Big Data' Quality Assurance

Achieving a high level of data quality is expensive. It is therefore normal for organizations that collect, store and use data to compromise quality by trading it off against cost. As a consequence, inappropriate uses of data can only be avoided if information is available about the data's quality and the quality assurance processes that have and have not been applied.

A framework for 'big data' quality assurance should involve:

- Incorporating applications of 'big data' within the organization's risk assessment and risk management framework;
- Incorporating applications of 'big data' within the organization's data quality assurance framework;
- Ensuring that the organization's data quality framework addresses the intrinsic and contextual data quality factors identified;
- Ensuring that data collections are not consolidated unless: they satisfy threshold data quality tests; their purposes, their quality and the meanings of relevant data-items are compatible; and relevant legal, moral and public policy constraints are respected;
- Ensuring that, where sensitive data is involved, particularly personal data, anonymisation techniques are applied, and the data that is submitted to analysis is not re-identifiable;
- Ensuring that, where data scrubbing operations are undertaken: they are undertaken within the context of the organization's data quality assurance framework; they involve external reference-points, and are not limited to internal consistency checks; their accuracy and effectiveness are audited; and the results are not used for decision-making unless the audits demonstrate that the results satisfy threshold data quality tests;
- Ensuring that inferencing mechanisms are not relied upon to make decisions, unless the applicability of those mechanisms in respect of the data in question has been subjected to independent review and they have been found to be suitable; and
- Ensuring that, when 'big data' is applied to decision-making: the criteria of relevance, meaning, and transparency of decision mechanisms are all satisfied; the results are audited, including by testing against known instances; and the outcomes are subjected to post-implementation assessment, including through transparency arrangements and complaints mechanisms.

Source: Clarke, R. (2014) *Quality Factors in Big Data and Big Data Analytics* (Revision 17th September 2014). http://www.rogerclarke.com/EC/BDQF.html

Nevertheless, the costs of production, curation and sharing must be met. Data may be free, but collecting, curating, hosting and making it available costs money. Nor is there a lesser role for the market. For example, openly accessible, quality, authoritative and trusted hydrological data (e.g. water depth, flow rate, rainfall, etc.) enables and underpins the market for water allocations throughout Australia's Murray-Darling Basin. A market vital for the efficient allocation of a scarce resource (water).

Instead of asking who will pay, a better question might be: how can we find a way to pay that does not raise access barriers, stifle innovation, and reduce welfare?

5. Temporal analysis and value

There are major difficulties in measuring the value and impact of data as anything but a snapshot, as they can change over time. This section explores how and why value and impact might change.

5.1 Changes in value

There are many reasons for the value of existing data to increase over time, or for their value to fluctuate from time-to-time, without any change or enhancement to the original data. For example:

- Time series data based on repeated observations, such as a household survey or national census, can often see a spike in use on initial release of the new data *and* in the use of previous editions on release of a new edition;
- New research techniques, new theories and new questions can lead to renewed interest in, and use of, data, and bring renewed/additional value to existing and older data collections;
- Similarly, new commercial uses of data and new products based on the data can lead to an increase use of the data, and bring additional value to existing and older data collections; and
- Anecdotally, I have heard that a researcher's data can see a second burst of use after they retire or die – perhaps when colleagues are more comfortable about criticising the data or more comfortable about using the data because the person has finished with it, perhaps because they are then confident that the data creator will not accuse them of misinterpreting or misusing the data, or perhaps because people want to carry the work on and seek to re-visit and re-familiarise themselves with the work.

Studies of established UK research data centres suggest that the value of the collection can increase over time, independent of any change in the value of the individual elements within the collection. For example, there was some evidence from the study of the UK Archaeology Data Service (ADS) that the value to users was increasing as the data collection expanded and the data service matured (Beagrie and Houghton 2013). What is not clear is whether, and to what extent, the apparent change in value was due to aggregation, extension and/or intensification of use, curation and access services improvements, or any number of other factors that might be at play. Nor is it clear, methodologically, how one might unpack these value changing factors.

5.2 Changes in impact

There is also the issue of impact, especially in regard to evidence and arguments influencing decisions. Policy decisions, as well as investment and business decisions, are binary: analysis can influence a decision in support of OR against doing something.

For example, there is a, probably apocryphal, story from the Research Councils UK (RCUK) quest to measure research impact, in which a Professor of European History is said to have

claimed "my research has contributed to the avoidance of World War III". It is plausible, and WWIII would have come at huge economic and social cost, so it is significant. But what if the work is merely delaying WWIII, until weapon systems are more expensive, lethal and destructive? That would turn a major positive impact into a negative impact, and is a reminder that apparently positive impacts can become negative over time, or turn out to be negative in hindsight. There are many examples, such as thalidomide.

Box 2 The issue of the counterfactual

The value or impact is the difference between what does happen and the next least worst option. So it is important to consider the counterfactual, both immediate and overtime.

In the UK data centre studies (Beagrie and Houghton 2012, 2013a, 2013b, 2014), tried to take some account of the counterfactual by asking users if they could have collected/created that data they used themselves or could have obtained it from elsewhere. Only the use by those who could neither collect/create the data themselves nor get it elsewhere is additional use, so only it should be measured – as additionality.

With some data, such as research data, it is often the case that it can only be collected/collected once, so there is clear additionality in additional uses. For example, it is often said that "all archaeology is destruction", as one can only excavate a site once. It cannot be repeated. Similarly, a national statistics agency can do a household survey or a census every 10 years, but one cannot do the 1979 household survey again. But this does not apply to all cases, and may not be typical of geo-spatial data.

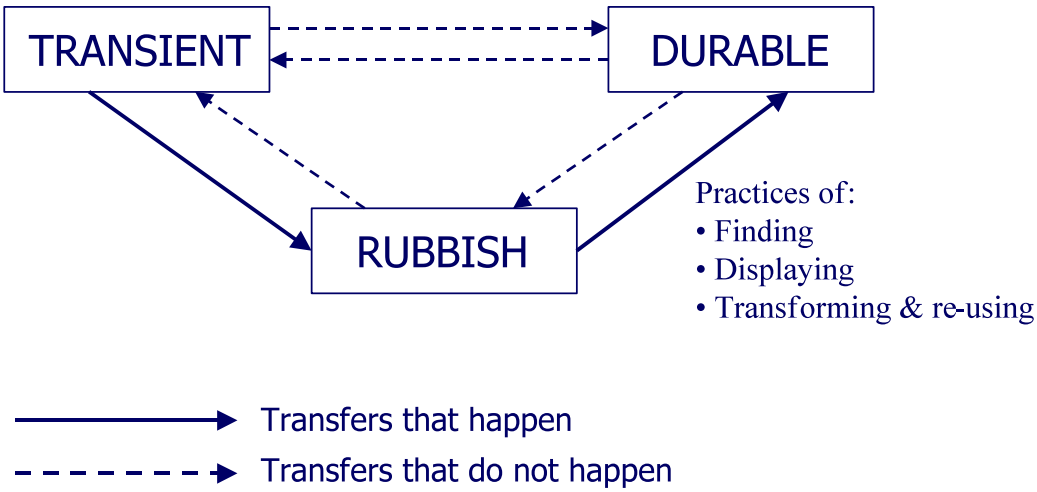
Nevertheless, studies such as the UK Ordinance Survey OpenData Impact Study by ConsultingWhere and ACIL Tasman (Carpenter and Watts 2013), using general equilibrium models, can include analysis of the base case of focus and the counterfactual in the form of the situation without the base case. While there are many uncertainties, this can enable estimates of value and impacts over a longer time period.

5.3 From transient to durable via rubbish?

In the context of changes in value over time, 'Rubbish Theory' (Thompson 1979) may be worth revisiting, because of its temporal angle regarding the transitioning of transients to durables via rubbish – suggesting widely fluctuating value (Figure 2). Examples of the cycle of value noted by Thompson include such things as antique furniture, vintage and classic motor cars, which are purchased and used, depreciate in value, are scrapped, then rediscovered and reused, and appreciate in value as a result of that re-use.[15]

15 Vintage and classic motor cars can be worth more in original condition than are restored vehicles, so it is not about the value-adding of restoration. http://www.hagerty.com/articles-videos/Articles/2009/06/02/Preservation-Class-Cars

Figure 2 Thompson's 'Rubbish Theory'



Source: Based on Parsons, L. (2008) 'Thompson's Rubbish Theory: Exploring the Practices of Value Creation,' *European Advances in Consumer Research*, Vol.8, pp390-393.

As Parsons (2008) suggests:

> *The focus on Rubbish Theory highlights the fact that much of value creation occurs beyond the first flush purchase of the new in the subsequent (re)uses, display and exchange of objects. Thus it focuses attention away from the moment of purchase and towards the ways in which objects are absorbed into our lives through cycles of (re)use. It also highlights the creativity of social actors in creating the conditions for value to emerge…*
>
> *For Thompson, rubbish can only really be understood in relation to the categories of transient and durable. Indeed these two categories represent the visible and valued elements of material culture as opposed to the invisible and unvalued 'rubbish'. It is important to note that these 'categories' ideally represent ways of seeing objects as opposed to substantive containers for them. The transient represents the usual state of commodities as objects which are declining in value and which have finite life spans. Whereas the durable increase in value over time and have (ideally) infinite life spans…*
>
> [For Thompson, 'The Practices of Value Creation'] *include: finding objects, displaying objects and transforming and re-using objects. It is argued that each of these sets of practices changes the way we view an object moving it from being seen as a 'rubbish object' of no value to a 'durable object' of increasing value…*
>
> [Thompson's work] *highlights the importance of thinking in terms of movement, flow and circulation and moves us away from means-end, supply-demand, production-consumption linearities in thinking through the consumption process… it suggests that value, rather than being an inbuilt property of an object, emerges through our ways of seeing and placing objects.*

There is much in Thompson's 'Rubbish Theory' that resonates today, including: (i) the idea that point-in-time estimates and measures may not fully capture value, (ii) that value can change in unforeseen ways, (iii) in making more shapely the distinction between use and re-use as quite different phases, and (iv) highlighting that the potential value of re-use is very difficult to predict.

6. Skills and capacity

The value and impact of data depends on the nature and extent of its use, and a crucial element are the skills and capacity of people to analyse and apply data throughout the economy and society. The fact that data scientists are often called 'unicorns', because the combination of skills required is so rare, and the kinds of salaries offered and reported, suggest that demand currently exceeds supply (Dwoskin 2014). There are reports of universities moving to modify and/or create new courses to develop suitable recruits for data scientist jobs (e.g. University of Virginia, Columbia University, and Ohio State), but it seems likely that much more could be done on the training side. Although, of course, there are 'training the trainer' and 'retaining the trainer' issues, which are always difficult when there are highly paid jobs widely available. Nevertheless, there are parallels throughout the 'ICT revolution' from which to learn.

Automation can be an important contributor as a (semi)substitute for skilled labour, on both the production and use sides. Inter alia, this means quality and metadata built-in on the original data creation/collection (production) side, and automated extraction and consumption (e.g. via APIs) on the user (consumption) side. There may also be parallels and lessons from the old debate about how much people needed to know about computers to be able to use them. Although there is also the issue of users (not) understanding the data enough to use it appropriately.

7. Summary and conclusions

Examining the open data policy context we suggest that, while countries confront their own particular issues and difficulties in going from policy to practice and progress is uneven, there is a trend towards more open data – at zero or marginal cost, with unrestrictive licensing, and machine readable in standardised data formats. There is also a shift towards being more inclusive, especially in the increasing expansion into static content and the GLAM sector (i.e. galleries, libraries, archives, and museums). We also note that there may be a wider open data opportunity, based around establishing a regulatory and policy environment conducive to enabling the maximization of opportunities to open private data.

Turning to the impacts of increased access, we note that the evidence suggests that providing open access to data leads to more use and creates greater value than do priced and/or restrictive access regimes. The principal dimensions of impact arising from increased access include:

- *Transaction cost savings* for both data providers and users, to the extent that direct agency cost savings alone could be sufficient to cover the possible loss of revenue arising from open access;

- *Efficiency impacts*, primarily for data users, but also for suppliers, that can be quantified as time-cost savings. In some circumstances, these savings may be reinvested in the activity, thereby having further return on investment impacts (e.g. where research time saving results in more research being done, thereby increasing the return on investment in research);
- *Additional use* that would not otherwise have been possible, thereby increasing the return on investment in the data collection/creation (e.g. pure additional uses of data by those who could neither create/collect the data themselves nor obtain it from elsewhere);
- Potentially significant *wider impacts* that are more difficult, perhaps impossible, to measure (e.g. greater engagement and safety during outdoor activities as a result of combinations of location-specific information and geo-tracking); and
- *New applications and combinations*, with unforeseen value and impacts emerging over time.

These impacts can be seen across data types and sectors.

Exploring new modes of consumption, we discuss the impacts of online delivery of digital content, the increasing length and complexity of value chains, and the expanding 'shared base'. Digitalisation makes content-based goods even more like public goods, in terms of both being non-rivalrous in consumption and non-excludable, bringing challenges as well as enormous opportunities. Noting that data are typically an intermediate good, with many and multiple possible uses and combinations of data feeding into many possible production processes, planned and serendipitous, we suggest that market signals are becoming more varied, fragmented, mediated, etc., and much less clear. Moreover, we suggest that the theoretical underpinnings of the private versus public trade-off shifts as the economy moves toward a digital one, with a larger role for government in the digital economy.

Examining the question "who will pay?", we suggest that the benefits of open access drive the need to resist too early commercialisation, where that involves placing restrictions on access and use. Exploring the alternative question "who will not pay, and why?", we note that innovative, integrative uses for which the outcome is uncertain (e.g. no established market) or is itself a public good (e.g. research and education, data/information), may well be excluded. Rather than focusing on the free-rider problem, we note that there are examples of commercial exploitation of the free-rider opportunity. Hence, a better question might be: "how can we find a way to pay that does not raise access barriers, stifle innovation, and reduce welfare?"

Exploring changes in value and impact over time, from both evidential and theoretical perspectives, we note many challenges facing those seeking to estimate the value of data and suggest that there is much in Thompson's 'Rubbish Theory' that resonates in this era of potentially unlimited use and re-use of data – including: (i) the idea that point-in-time estimates and measures may not fully capture value, (ii) that value can change in unforeseen ways, (iii) in making more shapely the distinction between use and re-use as quite different phases, and (iv) highlighting that the potential value of re-use is very difficult to predict. It also brings value perceptions to the fore, reminding us that we should not dismiss methods focusing on

perceptions (e.g. contingent valuation), and suggests that we need to consider the value of future ‘options’.

References

ABS [Australian Bureau of Statistics] (2013) *Annual Report, 2012-13*, Cat No 1001.0, Canberra. http://www.abs.gov.au/

Appleyard, G. (2014) 'Should PSMA data be available for a fee – or for free?' Locate14 Conference, Canberra Convention Centre, 7-9 April. http://www.psma.com.au/psma/wp-content/uploads/Glenn-Appleyard-Locate14-Keynote.pdf

Beagrie, N., Chruszcz, J. and Lavoie, B. (2008) *Keeping Research Data Safe*, JISC, London and Bristol. http://www.beagrie.com/krds.php

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

Beagrie, N., Houghton, J.W., Palaiologk, A. and Williams, P. (2012) *Economic Evaluation of Research Data Infrastructure* (ESDS), Economic and Social Research Council, London. http://www.esrc.ac.uk/_images/ESDS_Economic_Impact_Evaluation_tcm8-22229.pdf

Beagrie, N. and Houghton, J.W. (2013a) *The Value and Impact of the Archaeology Data Services: A Study and Methods for Enhancing Sustainability*, Joint Information Systems Committee, Bristol and London. http://www.jisc.ac.uk/whatwedo/programmes/preservation/ADSImpact.aspx

Beagrie, N. and Houghton, J.W. (2013b) *The Value and Impact of the British Atmospheric Data Centre*, Joint Information Systems Committee and the Natural Environment Research Council UK, Bristol and London. http://www.jisc.ac.uk/whatwedo/programmes/di_directions/strategicdirections/badc.aspx

Beagrie, N. and Houghton, J.W. (2014) *The Value and Impact of Data Sharing and Curation: A Synthesis of Three Recent Studies of UK Research Data Centres*, Joint Information Systems Committee (Jisc), Bristol and London. http://repository.jisc.ac.uk/5568/

Cabinet Office (2013) *Policy paper: G8 Open Data Charter and Technical Annex*. https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex

Carpenter, J. and Watts, P. (2013) *Assessing the Value of OS OpenData to the Economy of Great Britain: Synopsis*. https://www.gov.uk/government/publications/ordnance-survey-open-data-economic-value-study

Chesbrough, H. (2003) *Open Innovation: The new imperative for creating and profiting from technology*, Harvard Business Scholl Press, Boston MA.

Chesbrough, H., Vanhaverbeke, W. and West, J. (eds.) (2006) *Open Innovation: Researching a New Paradigm*, Oxford University Press, Oxford.

Davis, P. M., Lewenstein, B. V., Simon, D. H., Booth, J. G., & Connolly, M. J. L.(2008) Open access publishing, article downloads, and citations: randomised controlled trial, *British Medical Journal,* 337(7665), 568.

DotEcon (2006) *The commercial use of public information (CUPI)*, Report oft861, Office of Fair Trading, London. http://www.opsi.gov.uk/advice/poi/oft-cupi.pdf

DTLR (2002) *Economic Valuation with Stated Preference Techniques*, London: Department of Transport, Local Government and the Regions. http://www.communities.gov.uk/documents/corporate/pdf/146871.pdf

Dwoskin, E. (2014) Big Data's High-Priests of Algorithms: 'Data Scientists' Meld Statistics and Software for Find Lucrative High-Tech Jobs, *The Wall Street Journal*, 8 August 2014. http://online.wsj.com/articles/academic-researchers-find-lucrative-work-as-big-data-scientists-1407543088

Gans, J. (2012) *Information Wants to Be Shared*, Harvard Business Review Press, Boston MA.

Gruen, N. (2010) *Mr Gruen goes to Washington (again)*, Club Troppo (Blog). http://clubtroppo.com.au/2010/09/16/mr-gruen-goes-to-washington-again-2/

Gruen, N. (2014) *Driving Government innovation without money*, Presentation at the Department of Communications, Canberra. https://www.youtube.com/watch?v=7YmD1gg7txY

Houghton, J.W. (2006) *Global Chains: Australia's challenge in the evolving world economy*, Competing from Australia Paper No 1, Committee for Economic Development of Australia (CEDA), Melbourne. http://ceda.com.au/public/publications/info_paper/cfa1.html

Houghton, J.W. (2009a) *Costs and Benefits of Alternative Publishing Models*: *Denmark*, DEFF, Copenhagen. http://www.knowledge-exchange.info/Admin/Public/DWSDownload.aspx?File=%2FFiles%2FFiler%2Fdownloads%2FDK_Costs_and_benefits_of_alternative_publishing_models.pdf

Houghton, J.W. (2009b) 'Exploring the impacts of enhanced access to publicly funded research,' in *The socioeconomic effects of public sector information on digital networks: Towards a better understanding of different access and reuse policies,* Ed. Paul Uhlir, National Academies, National Academies Press, Washington DC. (June 2009). http://books.nap.edu/openbook.php?record_id=12687&utm_source=WIDa691682009071534634&utm_medium=Widgetv3&utm_content=12687&utm_campaign=Widget&utm_term=homeview

Houghton, J.W. (2009c) *Open Access: What are the economic benefits? A comparison of the United Kingdom, Netherlands and Denmark*, Knowledge Exchange, Brussels. http://knowledge-exchange.info/Default.aspx?ID=316

Houghton, J.W., and Sheehan, P. (2009) Estimating the potential impacts of open access to research findings, *Economic Analysis and Policy* 39(1). http://www.eap-journal.com/vol_39_iss_1.php.

Houghton, J.W., de Jonge, J. and van Oploo, M. (2009) *Costs and Benefits of Research Communication: The Dutch Situation, S*URFfoudation, Utrecht(June 2009). http://www.surffoundation.nl/wiki/display/economicstudyOA/Home

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). http://www.jisc.ac.uk/publications/publications/economicpublishingmodelsfinalreport

Houghton, J.W., Rasmussen, B. and Sheehan, P. (2010) *Economic and Social Returns on Investment in Open Archiving Publicly Funded Research Outputs*, The Scholarly Publishing and Academic Resources Coalition (SPARC), Washington DC. http://www.arl.org/sparc/publications/papers/vuFRPAA/index.shtm

Houghton, J.W., Swan, A. and Brown, S. (2011) *Access to Research and Technical Information in Denmark*, Forsknings- og Innovationsstyrelsen and DEFF. http://fivu.dk/en/publications/2011/access-to-research-and-technical-information-in-denmark

Houghton, J.W. (2011) *Costs and benefits of data provision,* Report to The Australian National Data Service, Canberra. http://ands.org.au/resource/cost-benefit.html

Houghton, J.W. and Gruen, N. (2012) *Excepting the Future: Internet intermediary activities and the case for flexible copyright exceptions and extended safe harbour provisions*, A Lateral Economics Report for The Australian Digital Alliance (ADA). http://www.digital.org.au/content/LateralEconomicsReports

Houghton, J.W. and Gruen, N. (2012) *Exceptional Industries: The economic contribution to Australia of industries relying on limitations and exceptions to copyright*, A Lateral Economics Report for The Australian Digital Alliance (ADA). http://www.digital.org.au/content/LateralEconomicsReports

Houghton, J.W., Dugall, B., Bernius, S. Kronung, J. and Konig, W. (2012) *General cost analysis for scholarly communication in Germany,* A report to the Deutsche Forschungsgemeinschaft (DfG) by the Goethe Universitat, Frankfurt. http://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/27530

Houghton, J.W. and Swan, A. (2013) Planting the Green Seeds for a Golden Harvest: Comments and Clarifications on "Going for Gold" *D-Lib Magazine 19(1/2) January/February 2013.* http://www.dlib.org/dlib/january13/houghton/01houghton.html

Kaplinsky, R. (2000) 'Spreading the gains from globalisation: what can be learned from value chain analysis?' *Journal of Development Studies* 37(2), pp117-146.

Lateral Economics (2014) *Open for Business: How Open Data Can Help Achieve the G20 Growth Target*, A Lateral Economics report commissioned by Omidyar Network. http://www.omidyar.com/sites/default/files/file_archive/insights/ON%20Report_061114_FNL.pdf

Nilsen, K. (2007) *Economic Theory as it Applies to Statistics Canada: A Review of the Literature*, Statistics Canada, Ottowa. http://www.epsiplus.net/media/files/nilsen_economics_paper_2007_final_version

Nilsen, K. (2009) 'Enhancing Access to Government Information: Economic Theory as it Applies to Statistics Canada,' in Paul Uhlir (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. http://www.nap.edu/openbook.php?record_id=12687&page=40

OECD (2005) *OECD Handbook on Economic Globalisation Indicators,* OECD, Paris.

OECD (2007) *OECD Principles and Guidelines for Access to Research Data from Public Funding*, Paris. http://www.oecd.org/science/sci-tech/38500813.pdf

OECD (2008) *Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information*, C(2008)36, Paris. http://www.oecd.org/internet/ieconomy/44384673.pdf

Parsons, L. (2008) 'Thompson's Rubbish Theory: Exploring the Practices of Value Creation,' *European Advances in Consumer Research*, Vol.8, pp390-393. http://www.acrwebsite.org/search/view-conference-proceedings.aspx?Id=13952

PIRA (2000) *Commercial exploitation of Europe's public sector information*, European Commission, Brussels. ftp://ftp.cordis.europa.eu/pub/econtent/docs/2000_1558_en.pdf

Pollack, R. (2009) *The Economics of Public Sector Information*, CWPE 0920, University of Cambridge, Cambridge. http://www.econ.cam.ac.uk/dae/repec/cam/pdf/cwpe0920.pdf

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. http://ijsdir.jrc.ec.europa.eu/index.php/ijsdir/issue/view/20

Quiggin, J. (2010) *The value of Public Sector Information for Cultural Institutions*, Report to the Gov 2.0 Taskforce. http://gov2.net.au/projects/project-6/

RIN [Research Information Network] (2014) *Nature Communications Citation Analysis*. http://www.nature.com/press_releases/ncomms-report2014.pdf

Science-Metrix (2013) *Proportion of Open Access Peer-Reviewed Papers at the European and World Levels, 2004-2011*. Produced for the European Commission DG Research & Innovation (RTD-B6-PP-2011-2: Study to develop a set of indicators to measure open access). http://www.science-metrix.com/pdf/SM_EC_OA_Availability_2004-2011.pdf

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. http://www.ist-world.org/ResultPublicationDetails.aspx?ResultPublicationId=5b3ae433c24747f489e0ad4470aa810b

Swan, A. (2010) *The Open Access citation advantage: Studies and results to date.* http://eprints.soton.ac.uk/268516/

Swan, A. and Houghton, J.W. (2012) *Going for Gold? The costs and benefits of Gold Open Access for UK research institutions: Further economic modelling*. Report to the UK Open Access Implementation Group. http://ie-repository.jisc.ac.uk/610/

The Danish Government / Local Government Denmark (2012) *Good Basic Data For Everyone – A Driver for Growth and Efficiency*, Copenhagen, October 2012. http://uk.fm.dk/publications/2012/good-basic-data-for-everyone/~/media/Publikationer/Imported/2012/Gode%20grunddata%20til%20alle/Basic Data_UK_web_2012.10.08.ashx

Thompson, M. (1979) *Rubbish Theory*, Oxford University Press.

Thorpe, J. (2003) *Economic perspectives on copyright law*, Centre for Copyright Studies, Sydney. http://www.copyright.com.au/assets/documents/research-reports/CCS0203Thorpe.pdf

US Department of Commerce (2014) *Fostering Innovation, Creating Jobs, Driving Better Decisions: The Value of Government Data*, Department of Commerce, Washington http://www.esa.doc.gov/Reports/fostering-innovation-creating-jobs-driving-better-decisions-value-government-data

Vickery, G. (2010) *Review of recent studies on PSI re-use and related market developments*, European Commission, Brussels. http://ec.europa.eu/information_society/newsroom/cf//document.cfm?doc_id=1093

Vickery, G. (2014) *Public Sector Information: A review of the Recommendation*, DSTI/ICCP/IE (2012) 2/REV3, OECD, Paris.

Wagner, A. Ben (2010) Open Access Citation Advantage: An Annotated Bibliography, *Issues in Science and Technology Librarianship*. Vol 60. Winter 2010. http://www.istl.org/10-winter/article2.html

Ware, M. (2009) *Access by UK small and medium-sized enterprises to professional and academic literature,* Publishing Research Consortium, Bristol. http://www.publishingresearch.net/SMEaccess.htm

Weiss, P. (2002) *Borders in Cyberspace: Conflicting Public Sector Information Policies and their Economic Impacts*, US Department of Commerce, Washington. http://www.nws.noaa.gov/sp/Borders_report.pdf