Permission granted: The economic value of data assets under alternative policy regimes

A Lateral Economics report for the Open Data Institute

March 2016

LateralEconomics

CAPABLE, INNOVATIVE, RIGOROUS

Executive Summary

Introduction and background

The myriad and continually growing uses of data – public sector information (PSI) and other data (including research and private data) have great social and economic value. However, there are formidable challenges in estimating the economic value of much of this.

- First, the phenomenon of data is vast and multifarious. There are innumerable data series, all with specific micro-economic characteristics. Data is typically acquired by intermediaries and/or developers and distributed via a great many products and services.
- Second, it is extremely difficult to arrive at a measure of economic value (e.g. pound sterling) of the final consumption of data because so much of it occurs for free. Even if one successfully addresses the conceptual issues (for instance, by envisaging some consumer 'willingness to pay'), the practical challenges of obtaining empirical evidence remain. Free consumption is part of the increasingly important 'dark matter' of GDP.

As much as we might be tempted to think the quantification of such uncertain effects irresponsible, those who must make important decisions are entitled to press analysts for their best guess of the indicative magnitudes we could be dealing with.

As one would expect, given this level of ignorance and uncertainty, existing empirical estimates of the value of open data vary considerably in scope. However, they suggest that the value added associated with open data varies between 0.4 and 1.4 per cent of gross domestic product (GDP) with the wide margin between these two numbers providing some quantification of our ignorance.

ODI has asked Lateral Economics to assist it to consider the economic implications of the commercial terms on which core data sets which form the bulk of PSI (e.g. land registry data and transport data) are distributed.

Pricing and licencing

Data providers incur various costs in acquiring, curating and distributing data. They may attempt to recover some or all costs by charging for access to the data. Or they may go further and maximise their own financial return. At the opposite end of this spectrum, data can be provided free and open licenced.

In many industries, cost-reflective pricing is efficient. However, with information goods like data, once it is made publicly available, the marginal cost of additional distribution is effectively zero. Thus, pricing at above this point will reduce demand and so curtail some information distribution that would be cost effective. On the other hand, just as private firms must find the

wherewithal to meet all their costs, so government agencies will sometimes find it appropriate to charge for data to meet fixed costs even though the marginal cost of *additional* distribution is effectively zero. Accordingly, this study investigates the magnitude of the economic effects of this latter course.

The impact of changing price regimes

To provide an indicative estimate of the impact of shifting from cost-recovery pricing to open data, we build on prior work that estimates the impact of reducing the prices of PSI. We estimate that the increase in re-use of data from removing licence restrictions is similar in magnitude to the impact of dropping prices to zero. In terms of the value created, a shift from a cost-recovery to an open-access regime is likely to more than double the value of the re-use of the data, adding around 0.5 per cent to GDP.

The impact of moving in the other direction – from an open-access regime to a cost-recovery regime – will reduce the impact, perhaps by around half, at least in the shorter term. This is because once the search for new and innovative uses has been done under open data, those subsequently charging for data have an interest in preserving that outcome. Nevertheless, once charging and licensing is introduced, the search for *further* beneficial uses for data will be curtailed which will see the loss from charging gradually climb back towards 0.5 per cent.

The implications of a profit-maximising regime are more uncertain. On the one hand, the revenue from sale of data is likely to rise – producing a further fall in demand, suggesting losses greater than 0.5 per cent. However, a sophisticated profit maximiser would probably do considerably less harm than might be expected from a firm that priced its data products crudely. It would seek maximally open options to monetise its data – such as advertising and freemium access. Further, a savvy profit maximiser might invest in additional data collection, curation and quality assurance work to optimise the value of its product. However, our report identifies substantial risks in such a course.

Implications

Across the value chain (with the exception of the acquisition of core data), there appear to be no material barriers to competition. So we expect reductions in costs to make their way to the ultimate consumers – the public. There will be some exceptions (for example, where firms can enhance existing products) and there will be winners and losers where there are risky developments. Empirical studies suggest that once open data is provided, demand for re-use will rise rapidly. However, the full value to final consumers may take some time to eventuate as new applications are developed.

The two biggest obstacles to further developing the market for core data assets are, as they have been in the past, apathy and/or opposition within data providers to opening their data and investing in optimising its quality for *users* rather than solely the PSI producers. But the UK is a world leader in tackling these issues. And with further effort comes further opportunity.

Table of contents

1 Introduction

The mission of the Open Data Institute (ODI) is to 'connect, equip and inspire people around the world to innovate with data'. A cornerstone of achieving this mission is encouragement of open access to data, particularly that produced by government agencies. The ODI seeks to better understand the implications of charging for data and has engaged Lateral Economics to explore the economic implications of paid access compared to open licencing.

Lateral Economics has been asked these questions:

1) What is the expected economic value to a national economy from core data assets under two different access models?
   - **a)** Paid access – where all organisations must sign a contract, pay a fee and potentially abide by licence restrictions on the purposes that the data can be used for
   - **b)** Open licence – where anyone can access, use and share the data
2) Assess how this value will accrue across the different parts of the value chain that use the data to deliver products/services to the market.
3) What is driving any difference in economic value between the two access models, e.g. network effects, time saving, allocative efficiency, etc.?
4) What are the expected timescales to realise the change in economic value if data is moved between these different access models (for example, from paid to open, or from open to paid)?
5) What steps can be taken to accelerate the change in economic value?

The rest of this document addresses these questions:

- Chapter 2 provides background including a description of the value chain and recent estimates of the value of open data
- Chapter 3 explores the economics of paid and unpaid access in a qualitative way
- Chapter 4 offers an indicative quantification of those issues
- Chapter 5 offers comments on the remaining matters in our terms of reference
- Chapters 6 and 7 incorporate References and Appendices

2 Background

2.1 An overview of core data assets

The core data assets that are the subject of this study include data assets such as land registries, ordnance surveys, meteorological data and transport movements. This data includes data that is purposefully collected (e.g. national statistics and meteorological data) and data collected as a by-product or some other function (for instance, business registration and court records). These data sets are most commonly referred to as public sector information[1] (PSI) as they are almost always generated by (or for) public agencies, referred to in this report as PSI holders (PSIH).

Discussions on the re-use of data, commonly focus on PSI. However, there are other sources of data. Two other important sources are:

- research or science data, especially that arising from publicly funded research, and
- private sector data. For example, sales data collected by private sector firms may be useful for economic management.

There are a great many applications of such data. Coupled with growth in the growth of applications, there has been growing recognition that such data has great value.

The data can be used to add value in myriad areas of the economy in myriad ways. It:

- reduces costs in providing services both by government and private sector (i.e. doing the same for less cost)
- enables new services and improved quality of services, and
- improves accountability for government services indirectly improving responsiveness and performance and in the process engendering greater trust in government.

Many of the benefits accrue directly to consumers of products and services that have made use of the data. However, there are

[1] Note, in this report, except where otherwise made clear, the words “data” and “information” are used interchangeably.

also benefits that accrue to the wider community. These benefits include, for example:[2]

- public benefits associated with improved transparency of government
- improved social cohesion, and
- positive externalities that may arise. For example, one person's use of transport data to optimise transport usage can improve traffic management and reduce congestion for other users.

2.2 The value chain for re-use of data

There are many descriptions of the value chain for the re-use of core data assets.[3] For our purpose, we propose to describe the value chain in terms of the following groups:

1. Data providers – these include PSIH, other government organisations and private organisations and individuals
2. Intermediaries, including
    a. data aggregators who source data from existing open data sources into a more useable form
    b. data enablers who facilitate the supply or use of data through reorganisation and reformat
3. Providers of products and services to consumers including
    a. developers who create applications for individual consumption
    b. data users who use data to enhance existing products and services or create new ones
4. End users (being the ultimate beneficiaries), including:
    a. direct users of the end product, and
    b. others who indirectly benefit from the open data usage by direct users.

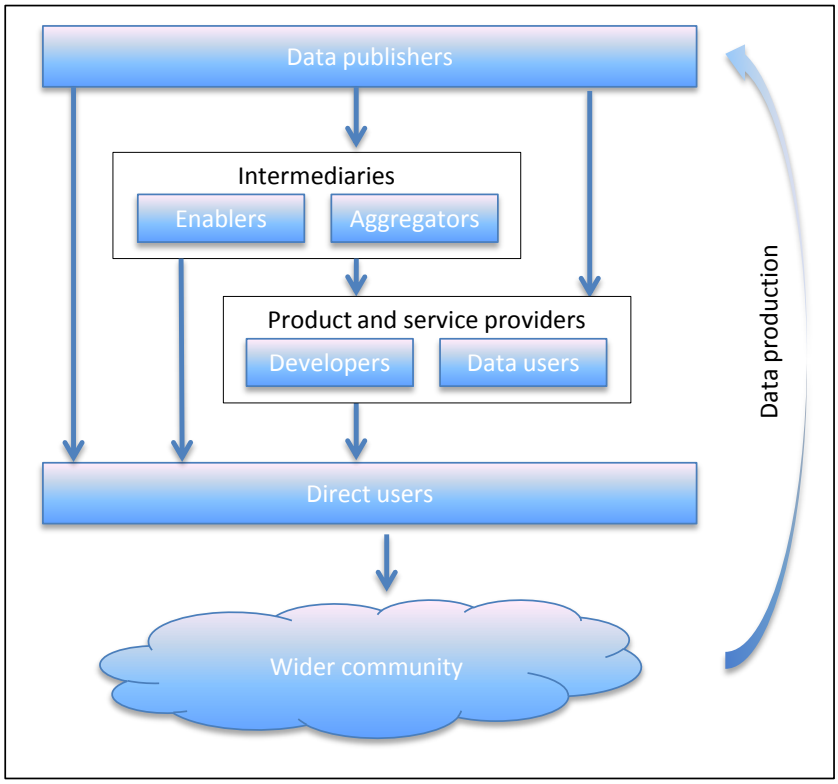
The connection between the groups is depicted in Figure 1 below.

---

2 This list is adapted from Deloitte (2013, p. 182).

3 See, for example, Deloitte (2013, p. 85), POPSI (2011, p. 14), PIRA (2000, p. 14) and Vickery (2011, p. 13).

Figure 1: Value chain for core data assets re-use



Source: Lateral Economics.[4]

The market characteristics of each group are discussed in Section 3.3.5. However, we note here that publishers or core data-sets are typically sole suppliers of information because they have privileged (often statutory) access to information, there are sunk costs in collection, they receive government funding or because they have an established reputation for quality. As a result, data providers have (at least some) power in determining data prices.

Reviewing the pricing policy for PSI, a European Commission study (De Vries et al., 2011, pp. 25-30) finds it useful to

[4] Any of the stages of production may be conjoint with any other. Someone using Google Maps will often be providing data back to the app, for instance, on traffic flow. Data providers may deal with intermediaries or product and service providers and release data to consumers. Further, there may be multiple chains before reaching the end user. Thus, government data providers may share data with other agencies. Similarly, there may be multiple steps in the value chain before a product or service is provided to an end user. A 2006 survey of businesses regarding their use of PSI, found that less than 30 per cent of businesses used PSI to make products for consumers (OFT, 2006, pp. 28-29).

categorise (direct) end users as being high end or low end. It describes the high-end market as consisting of a relatively small number of re-users that provide high value-added services to meet the needs of professional clients.[5] In contrast, the low-end market consists of re-users providing low value-added services to a large number of clients (e.g. mashing up free content to integrate into other services). Such a distinction may be important for pricing policies, as the high-end market will be less price sensitive.

2.3 Estimating the value of open data

There is a literature of reasonable size – to which Lateral Economics has contributed – estimating the benefits of open data (defined as data that anyone can access, use and share). This study seeks to measure the contribution of open data to gross domestic product, or GDP (which is measured by summing gross value added, GVA, through the production chain). Though the limitations of GDP are well recognised,[6] it is a convenient common metric.

Further, measuring the impact of some policy requires comparing different reference points (or scenarios). For example, the 'current' value of open data might be considered as the value that would be lost should the relevant data not be available. However, this later, hypothetical, scenario is difficult to envisage. It leads to consideration of what substitutes the market might develop should existing data be inaccessible.

The period over which value is created is also relevant. We are interested in the future, but this may differ substantially from the past (which we can measure). Other measurement challenges include:

- there are a great many end uses of the products;
- there is very little data on the end use of the products; and
- the prices paid by consumers – very often zero – will generally be significantly less than the value derived.

---

5 De Vries *et al*. (2011, p. 25) give the example of a meteorological company that provides very detailed weather forecasts to oil rigs, based on enhancing PSI data.

6 More broadly, government policy should generally be concerned with wellbeing for citizens in the present and the future.

Thus we risk significantly, perhaps hugely, underestimating the value of data if we derive values from observed prices.

2.3.1 Empirical studies

Economic evaluations of the impact of open data may focus on particular applications or, as in this study, the whole economy. Similar economy-wide evaluations have varied in scope. Notable dimensions include:

- The sources of data considered. Most studies have focussed only on PSI. Other studies also consider the value of research data and private sector data.
- The region considered. Most studies have been limited to a single country (e.g. the UK) or a select group of countries (e.g. the G20).
- Sectors considered. Some studies have focussed on a specific sector (e.g. transport).
- The scope of benefits considered, in particular, whether wider benefits (such as relating to reduced corruption) are included.
- Whether the value considered is an existing value or some measure of potential value. Some studies have just focussed on the net benefits; that is the value added less the cost of production.

In Table 1 below, we summarise estimates derived from the results of some key studies. To facilitate comparison, for some studies we have applied some additional analysis to the results presented in the study (see footnotes to the table) and for consistency and convenience we express these as a percentage of GDP.

The table is divided into two sections. The top section describes studies that have attempted to measure the current contribution of open data. The bottom section refers to two recent studies that attempt to estimate the potential additional contribution of open data.

The approaches used vary. The approach used by PIRA (2000) has been described as "top-down" as it begins with value added and then attempts to assess the contribution of data to the value. OFT (2006) contrasted this with a bottom-up approach which is based on attempting to estimate the value derived (interpreted in terms of willingness-to-pay) by consumers.[7]

We expect the values of open data (as a percentage of GDP) to increase over time due to the rapid expansion of new applications and the greater opportunities for re-use by consumers as a result of increased penetration of digital devices.

The narrowest scope scenario we might consider is the current net benefits to direct users. The widest scope is future gross value added from PSI and other data to direct users and the wider economy. For the purposes of evaluating the issue of paid access for core data sets, we propose the appropriate reference point is core data assets , broad in terms of beneficiaries, and incorporating the value added in the near term (i.e. more than is just presently realised).

Based on the existing studies, we concluded that the current GVA of core data assets to the economy is in the order of 0.4-1.4 per cent of GDP.

[7] OFT (2006, p. 114) raised the concern that the top-down approach risks overestimation in part because it takes no account of the possible use of alternative data sources that might be used. Conversely, a risk of the bottom-up approach is underestimation, not least because of difficulties in accounting for wider impacts.

Table 1: Estimates of the value of open data

Study (year)	Country/ region	Measure estimated	As % of GDP	GVA % of GDP
Studies estimating current value				
PIRA (2000)	PSI in Europe	Total value added of PSI		1.4%
DotEcon (2006)	UK	Net surplus (i.e. net of costs of supply) of PSI, excluding wider economic benefits	0.25%	n/a
MEPSIR study (2006)	EU25 + Norway	Market size for PSI, excl. wider economic benefits	0.25%	n/a
Pollock (2011)	UK	Welfare gains of opening up of PSI in 2006	0.11-0.13%	0.3-0.4%[8]
Vickery (2011)	Europe PSI	GVA of PSI in 2008 (incl. wider economic benefits)		1.2%[9]
Deloitte (2013)	UK PSI	GVA of PSI (incl. wider economic benefits)		0.4%[10]
Studies estimating potential benefits				
McKinsey (2013)	Global	Potential additional value	1.4%[11]	n/a
Lateral Economics (2014)	G20 countries	Potential additional value from selected case studies	1%[12]	n/a

Source: Lateral Economics analysis. See Table 2 in the appendix for further details. A summary of many of the studies is provided at Lateral Economics (2014, sections 3.2 and 3.3).

8 Extrapolated from results and parameters of median estimate reported in Pollock (2011). The author had estimated the welfare gain of moving to free and open-access. We applied the core assumption and parameter values to estimate the value of open data to GDP.

9 Vickery's method was based on extrapolating estimates from previous studies.

10 We are not clear as to how this estimate was derived.

11 See Lateral Economics (2014, p. 30). This is the estimated contribution to cumulative GDP growth over the next five years.

12 See Lateral Economics (2014, section 3.5).

3 Implications of a paid-access regime

This chapter examines the contrasting economic implications of paid and open access to data. It begins with a description of the alternative regimes. To assess the economic impact of a paid-access regime relative to an open-data regime, one must first identify and describe the implications of paid access relative to open data. Subsequent sections explore specific aspects of the economics of data.

3.1 Charging regimes for data

This review compares paid- and open-access regimes. It is useful to distinguish between multiple models including:[13]

- Paid access
    - profit maximisation — whereby the data provider sets prices to maximise its profit
    - cost-recovery of data production — pricing to recover the costs of data production
    - cost-recovery of initially establishing data distribution for re-use
    - marginal cost pricing of additional distribution — setting a price equal to the cost of supplying data to an extra user, which for digital data is essentially zero, and
- Zero-priced access — where data is not charged for but subject to restrictions on its use and redistribution.
- Open data — data that anyone can access, use and share.[14]

Generally, we would expect the prices charged and the revenues raised to be lower as we move down the list above.[15]

---

13 This list is adapted from Pollock (2008, p. 8). Pollock also notes that many PSIHs also have the ability to charge those parties providing updates to the information. For example, PSIHs dealing with registration of property, vehicle and company ownership may fund their data collection and processing activities from those registering the item.

14 See http://theodi.org/blog/closed-shared-open-data-whats-in-a-name.

15 A risk with subsidised supply is that there will be an excessive supply (e.g. investment in the provision of data). Conversely, there is the risk that there will be insufficient incentive for investment. However, there will be exceptions, particularly where demand is very responsive to price. De Vries et al. (2011) noted cases where the lowering of prices resulted in increased revenues. Furthermore, as discussed below, a profit maximising data provider might employ a mix of pricing strategies. Another argument for cost recovery is that

3.2 Implications of different access regimes

To assess the implications of different access regimes, we first consider the common rationale for paid access before considering issues particular to open data.

3.2.1 The common rationale for paid access

In most markets, users pay, and this is highly efficient. User charges ensure that resources needed to provide services are taken up only where their value is demonstrated by users' willingness to pay. In this way, data consumers decide whether or not to fund the collection and dissemination of that data. In terms of the charging schemes discussed above, the argument often leads to an incremental cost charging regime, whereby the supplier (in this instance, the data provider) recovers the incremental cost of providing access.[16] A pricing regime can, therefore, send a signal as to the optimal investment.[17]

3.2.2 'Abundance thinking': The economics of information goods

However, while the fixed costs of data provision is often large, marginal costs of additional dissemination are often negligible. If prices are set above marginal cost, then additional use will be discouraged even though its benefits exceed the (negligible) additional costs. As in many other industries, marginal cost pricing leads to an under-recovery of costs as there is no revenue to fund fixed costs. In essential services, such as energy and water, the problem may be addressed with two sets of charges corresponding to fixed and marginal costs. However, such a strategy is only effective when the fixed price does not deter access.[18] Outside government, very low marginal costs of information

---

the alternative suggests raising additional taxation revenue. This is costly because taxes generally distort behaviour away from what is socially optimal.

16 As is discussed in Section 3.2.2.1, profit-maximising pricing can also be efficient under the (albeit seemingly rare) conditions that the supplier is able to differentiate its charging so as to capture any consumer surplus.

17 With regard to open data, this is likely to be the greater risk. As noted by Pollock (2008, p. 13), requiring an organisation to charge at less than average cost can reduce the incentive for the organisation to develop new products.

18 Two-part pricing is also operative on the internet where internet service providers typically charge a fixed fee for access to the internet and nothing further for use (sometimes up to some cap on data usage).

provision has produced various business models. In most services for the mass market, resources are provided by means other than prices.[19]

3.2.2.1 Profit maximisation and price discrimination

A concern with paid access is that the data publisher will attempt to go beyond the recovery of incremental costs and instead maximise its profit[20] *Prima facie* this would increase prices and further reduce demand. However, there are additional considerations. Profit maximisation may stimulate demand by price discriminating; charging users different prices depending on their willingness to pay.

In principle, perfect price discrimination is as efficient as perfectly competitive pricing, but the informational and behavioural demands on the seller to bring this about are Herculean.[21] In practice, price discrimination is usually difficult and can result in additional waste. Suppliers of information goods adopt a number of strategies. These include:

- Discriminating using quality of product or service whereby a lower price is offered for lower quality products. Common strategies are:
    - Windowing – whereby the product is brought to market at different times in different formats. For example, films are released first in the cinemas and then at a later date onto other mediums.
    - Versioning – whereby the product is released with differing levels of quality.
    - Service modifications – whereby there are differences in the level of support.
- Bundling, whereby to obtain a product, consumers are required to purchase a bundled package.

---

19 Open-source software leverages the voluntary efforts of software users fixing bugs or adding features. Wikipedia uses philanthropy to run a platform that users volunteer their time on. Facebook, Twitter and Google provide their services for free while 'monetising' the social value they create from advertising revenue. Elsewhere a dominant strategy is 'freemium' – a form of price discrimination we explore further in the following section.

20 By this term, we do not mean marginal costs, but rather the full cost arising from the public distribution of the data. This may involve the full costs of all data collection, curation and distribution, or where governments already collect and/or curate the data in any event, all additional costs in curating and hosting the data for public release.

21 Thus each buyer's willingness to pay must be known and charged without anyone arbitraging the differences in prices in the downstream market.

Where lower quality products generate inconvenience or lower utility to the consumer, whilst saving the producer nothing, economic losses arise.[22]

Freemium

There is a growing army of products funded through 'freemium' business models in which data and other services are freely distributed whilst those seeking higher levels of quality or service pay. Free services are effective marketing including lowering buyers' transactions costs by allowing them to 'try before they buy'. Freemium products include LinkedIn, Google Apps, Evernote, Dropbox, Feedly, Pocket.

Chris Anderson (2009) coins the term 'abundance thinking' to describe the mindset that produces 'freemium' pricing – making the abundance of the digital world and its zero marginal cost of production – work for consumer and user alike. As Gruen (2015, p. 5) puts it, while funding the fixed costs of these services raises the free rider problem, the near-zero marginal costs of expansion mean that the free rider opportunity will often trump the free rider problem.

Given all this, it is likely that a sophisticated profit maximising data provider would do considerably less harm than might be expected if a firm were to price its data products crudely. It is even possible to imagine circumstances where profit maximisation would provide incentives to invest in additional data collection, curation and quality assurance work, and that this could increase economic welfare above the level that might be achieved by a relatively mediocre government agency administering government mandated policies of open data.[23]

---

22 Since the price discriminating firm's sole interest is to maximise its access to consumer surplus without regard to the resulting disutility to customers, it may do more harm to the total utility to consumers than its expansion of supply benefits economy-wide welfare. These issues are not new. The economist Jules Dupuit raised concerns in 1849 with regard to price discrimination of railway carriages with little left undone to make conditions unpleasant for third class passengers, not to save costs, but to avoid second class passengers buying third class tickets. Today's mobile phone packages may well provide a contemporary example of price discrimination which lowers general wellbeing given the additional costs of staying within plans and the cost of informing oneself of their respective terms. Nevertheless, the digital age gives us a new twist as the complexity of modern mobile phone plans also establishes a 'confusopoly' making it harder for consumers to understand various trade-offs.

23 Google Maps has invested substantial sums in generating and curating data for distribution which it can monetise by charging premium users. Nevertheless, it offers free access to a standard product for the vast bulk of direct users of the service. https://developers.google.com/maps/pricing-and-plans/#details

Nevertheless, it seems unlikely, and there are further risks in such a course. First, the entrepreneurial flair of a highly innovative profit maximising incumbent might give way to more complacent behaviour in the future in which a more mature firm's managers use their monopoly position to meet quarterly revenue and profit growth targets. Second, the data would almost certainly be distributed according to licencing restrictions which is likely to seriously curtail economic welfare (see Section 3.2.3 below). Third, if we can imagine excellence in harmonising general economic welfare with profit maximisation, we can surely imagine excellence in the public sector which can target economic welfare more directly without the additional imperative of profit maximisation with all the distortions it entails here.

3.2.3 Transactions costs

The transactions costs borne by consumers and the suppliers differ greatly between access regimes. The process of vending data is almost by definition more complex than simply disseminating it for free. The costs to the supplier include administrative costs such as invoicing, as well as costs of managing a licensing and compliance regime, and De Vries concluded they were significant (2011, p. 6). They include:

- Building an online sales environment where the qualities of data are described pending their sale.
- Building the relevant security layers or sub-contracted platform services to take commercial payment.
- Commissioning the work to know what kind of licensing terms to impose and then the legal work to design those terms.
- Considering whether or not to take action against those who breach them and, if so, funding that.

Nevertheless, there are potentially more profound forces at work. Information is a non-rival good. Use in one application does not preclude use in another. And with near zero costs of distribution, even small transactions costs can be a big deal as has been recently illustrated on the internet (See Box 1 below).

Box 1: The significance of transactions costs on the net

The global phone network and the internet are both built around 'interconnect' agreements in which nodes on the network exchange access to each others' users. The phone network facilitates dedicated connections between users. So large telcos' negotiate interconnect agreements fiercely, with each seeking to maximise its 'cut' of the economic rent.

The internet works by routing addressed data packets, each making its own opportunistic way through the net depending on network conditions. If someone won't negotiate interconnect reasonably, others can be found and, so, few are tempted to negotiate unreasonably. As a result, transactions costs between service providers negotiating reciprocal access to each others' services collapse. Virtually all - 99.5 per cent – of reciprocal access agreements occur informally without written contracts. What does this mean for efficiency and productivity? On an equivalent voice-per-minute rate, internet rates are around one hundred thousandth of typical voice rates.

The collapse of transactions costs in cyberspace has led to the burgeoning of new social and economic formations. Anyone – including (crucially) any innovator – can access the network without requiring the permission of, or paying rent to, monopolistic gatekeepers – as one must with telephone or TV networks.

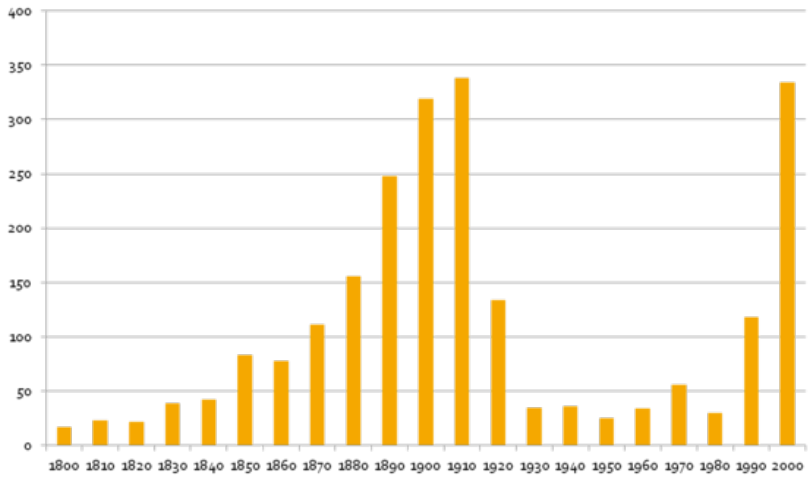
Adapted from Gruen (2012).

Prices impose transactions costs on users and, given that they are borne by *each user,* they constitute a potentially much larger source of deadweight loss. These costs not only include administrative costs such as those associated with reviewing licence agreements and making financial transactions but also — as described by Szabo (1999) — "mental transaction costs" to consumers. Szabo has categorised them into costs associated with dealing with uncertain cash flows, observing product attributes and complexity of decision making. If these are 'cognitive' costs, there are also 'psychological' transactions costs. As Chris Anderson has documented (2009), free is a very special price and, for many consumers, a quantum leap beneath very low prices. Free means free of the risk of losing money, free of being taken advantage of, free to trust or to suspect vendors after inspecting their goods.

The implications of transaction costs are demonstrated in the figure below, which documents the huge rise in the availability of book titles on the market once copyright expires. The paradox is that there is demand for books from which publishers and copyright owners could make *some* profit (as the sale of out of copyright books demonstrates),

which they nevertheless forego. In other words, in the absence of transactions costs, one would expect *more* book titles to be in print during the copyright term rather than less, because the copyright increases the potential profit in their sale. And the *magnitude* of the effect is large – with book titles reduced by over 80 per cent.

Figure 2: Book titles in print from Amazon warehouse by decade



Source: Rosen, 2012.

There are also other costs to consumers to consider. A significant risk to commercial users of open data is that future supply will no longer be available or its quality will fall or that access will become more limited. Such concerns are in effect a cost borne by consumers. In sum, the transactions costs associated with charging for data and/or licensing that data to control redistribution can be substantial, but largely disappear under an open-data regime.

3.2.4 Network externalities and innovations

For many information goods, and in particular for data assets, both the supply and consumption of data can stimulate greater demand for several reasons. First, there can be consumption externalities. Thus, for instance, people using real-time transport data to avoid congestion lower congestion for everyone. Second, the rate of adoption of a particular service may increase with the penetration of the market due to the social influence of early adopters on later adopters.[24]

[24] For example, persons who use a data-enabled app (e.g. that provides real-time transport) may do so because they were told about it by a friend and/or influenced in their decision to adopt it from observing others. Such effects are commonly discussed in the literature on 'diffusion of innovations'.

Third, there are network effects associated with different re-use. Great value can often be derived when data sets are integrated with one another; for instance, TripAdvisor adds to the accumulated value of geospatial data and customer ratings data to help people identify and find travel destinations to their taste. Greater value still could be generated if its maps also integrated with live transport data. The greater the number of data-sets accessible, the lower the technical and commercial barriers to their integration, the greater the value generated by *each* data series.

Finally, there are supplier network effects. Increasing the number of developers using a particular data-set can stimulate additional development through a number of mechanisms. Greater re-use can result in economies of scale in the provision of intermediary services (i.e. by aggregators and enablers). Perhaps more significantly, there are network benefits in terms of innovation as developers help each other out in developer communities. *Prima facie,* we might expect that a charge on information would not materially inhibit the development of an innovation where the benefits far exceed the information costs.

However, empirical evidence suggests that even a small charge may significantly impede innovation.[25] There are several reasons for this.[26] First, the returns to the innovation may be highly dispersed among suppliers. The parties purchasing the data may not expect to recoup their investment as most of the value is captured further down the value chain. Second, the cost of obtaining information (including the costs associated with licensing) may need to be borne by multiple parties involved in development. Third, sellers are unlikely to know of all the ways their data can be valuable to others, and the magnitude of that value and this uncertainty is likely to make negotiating access a fraught process as each party seeks to capture what it sees as its share of benefit.

The combination of the above effects have prompted a number of parties to argue that the priority strategy for information goods should be on abundance of use as this will in turn stimulate greater supply and demand.[27]

---

25 Pollock (2008, Appendix A2) notes that "*Weiss (2004) argued, marginal cost access to weather data in the US was a large factor in the development of the multi-billion dollar weather derivatives industry".*

26 Some of these are discussed by Pollock (2008, Appendix A2).

27 There appears to be broad support from researchers (e.g. Pollock 2011, Vickery 2011, and Shakespeare 2013) for open data. There is also public support. De Vries et al. (2011, pp. 10-12) note that the majority of responses from public consultation were in favour of free access.

4 Estimating the impact of a paid-access approach

4.1 Overview and approach

The discussion above highlights why, for several reasons, paid access results in sub-optimal re-use of data. To estimate the impact of paid-access, we have used the estimates from Chapter 2 on the value of open data as a baseline. In effect, we are estimating the economic value lost through paid access.

The impact of paid access relative to open data depends in part on which paid-access pricing approach is employed; whether, for instance, it focuses on profit maximisation rather than cost recovery. The next section considers methodological issues.

4.2 Basic approach

The difference between open access and paid access to data is illustrated in Figure 3 below. It shows the demand for core data assets and the 'effective price' paid under different pricing regimes. We have defined the 'effective price' as the financial cost plus the costs associated with complying with any licence agreements. Of note, this price under a free-but-restricted regime is greater than zero.[28] A shift from open data to 'free but restricted' will increase transactions costs for sellers (see above) and may also raise 'mental transactions costs' for consumers. It is also likely to depress indirect demand for the data by those who might have otherwise received the data through the initial customer, but did not because the customer was not authorised to pass it on.[29] The indicative shape of the demand curve in Figure 3 is consistent with the conclusions of De Vries et al. (2011, pp. 25-30) who argue that sufficient price reductions open up a large low-end market.

The supply curve (which reflects the marginal cost of supply) varies by regime. Under open access it is, in effect, the horizontal axis.[30] Another related effect not captured in the illustration is that under paid-access

---

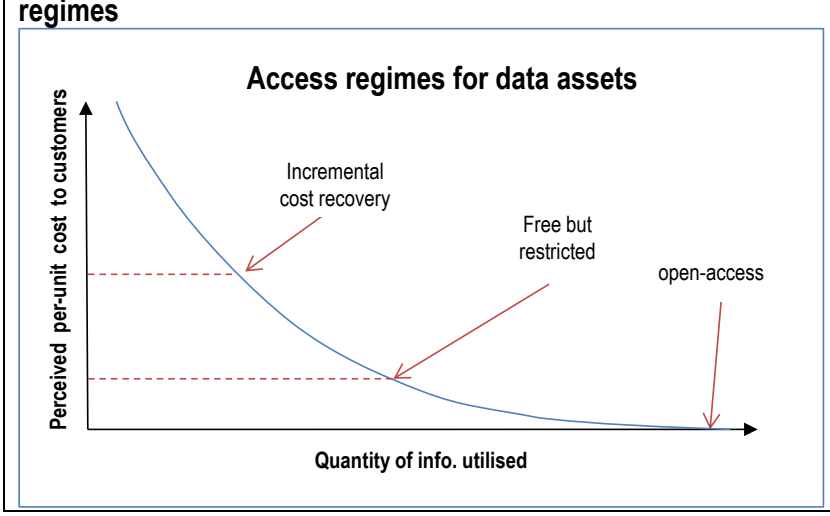
28 There are potentially alternative ways of illustrating this additional impact, (e.g. including a separate demand curve) however, this seemed the simplest approach.

29 Thus, for instance, if Hansard data was only available directly to users, but not for free redistribution, the organisation *They Work For You*, (http://www.theyworkforyou.com/) which substantially increases the distribution of Hansard information, might have been discouraged from distributing it.

30 More precisely the marginal cost of supply is high for the first consumer but effectively zero for each subsequent user.

regimes, the cost of supply also shifts upwards (from zero to a positive amount).

Figure 3: Demand for core data assets under different pricing regimes



A shift between charging regimes will be associated with a change in the value added and societal welfare. However, there are several problems with attempting to quantify these effects. Estimating the demand is particularly difficult. There is some information collected on demand and how this responds to changes in price;[31] however, this will not be representative of the welfare associated with open data.

As Pollock (2008) notes, there are two key issues. First, data is typically distributed to intermediaries and developers, not end consumers. The demand information captured, therefore, does not represent what the final consumers are willing to pay and the welfare gains to consumers. Because data can be re-used at negligible cost and developers are not able to capture many of the resulting consumer benefits, they are likely to underestimate them by a considerable amount. With much of the data supplied at zero price, there is no market signal of its value.[32]

Second, the information captured will represent the demand when it was captured, yet with the rapid change that characterises the area, the present may be a poor guide to the future. [33]

---

31 Pollock (2008) provides a useful summary.

32 For example, simple economic accounting for the value of Google would suggest that it is limited to its value as an advertiser, yet more sophisticated attempts to measure its economic value produce conservative estimates several times higher than this with debate ranging from ten to one hundred times the amount directly recorded in GDP. See Worstall (2015).

33 There are numerous additional issues in measurement. For example, the volume of direct access to a data-set may decrease as a result of consumers

Another set of issues relates to assessing the pricing regimes that might emerge. As we discussed in Section 3.2.2.1, an organisation might employ a variety of pricing strategies and business models, including approaches that simultaneously seek to maximise profit and re-use.

4.3 Modelling the effect of changing price regimes

In this section, we estimate the effect of a shift between paid and open-access pricing regimes. Our initial focus is on the change between cost recovery and open data. We then consider the implications of a profit-maximising pricing regime in which price-discrimination policies might be applied.

4.3.1.1 A model for estimating the impact of paid-access

A useful starting point is the work of Pollock (2008 & 2011). Considering a number of the limitations itemised above, Pollock developed a model to estimate the welfare effects of moving from average cost to marginal cost pricing for PSI as a function of.

- the fixed costs incurred in producing and maintaining the PSI
- the responsiveness of direct consumers (technically the price elasticity of direct demand), and
- a demand multiplier that reflects the difference between direct customers' willingness to pay and the total value provided to all final customers, many of whom have no direct relationship with the data provider.

Of course, the challenge with this approach is obtaining reasonable estimates for the key parameters. The fixed costs of providing data may be estimated with a reasonable degree of certainty; however, direct observation of the elasticity of demand and the demand multiplier is not possible. Pollock offers estimates of elasticity and the demand multiplier based on a review of evidence from several sources.

Pollock's model (summarised in **Box 2** on page 31 in the appendix) is reasonably intuitive. The more elastic (price responsive) the demand, and/or the greater the multiplier, the greater the loss from charging for data.

---

choosing to access the information via new applications developed by intermediaries. For example, all else being equal, the volume of *direct* users of meteorological data sets may fall as a result of the development of weather apps on smart-phones that access the data via intermediaries.

Using his model, Pollock (2011) estimated the welfare gains in the UK in 2011 from 'opening up' (i.e. moving to marginal cost pricing). His estimates ranged from 0.11 per cent to 0.4 per cent of GDP, around four to 11 times the cost of providing PSI.

A study on Danish address data (DECA 2010) provides one opportunity to test the estimate. The study estimated that the annual benefits of open address data were EUR 14 million at an annual ongoing cost of EUR 0.2 million. Other information in the report suggests the cost of providing the data was higher. Nevertheless, the case study provides a result that is above the higher range estimated by Pollock.[34]

4.3.1.2 Refining the model

Pollock estimated the change to a marginal cost pricing regime, under which he notes it would be "natural for the PSIH to make the data 'openly' available" (2008, p. 9). However, our interpretation is that the model and parameter estimates are more consistent with a reduction in pricing and not a removal of restrictions on use. In particular, Pollock assumed that the demand curve is linear (i.e. does not curve as illustrated in Figure 3) and uses evidence of elasticity estimates that included cases where prices were reduced but were not made free.

As discussed in section 4.2, we expect that transaction costs for uses of data are significant and that, as a result, demand will expand significantly when moving from a free-but-restricted regime to an open-data regime. To account for this, we extend Pollock's model to include a kink when transaction costs are removed (see appendix 1 for details). This approach brings new challenges. As discussed below, there is some anecdotal evidence on the increase in direct demand when shifting an open-access regime is introduced.

However, we must place a value on that additional demand. There are a number of considerations. In standard economic models, all economic agents have perfect knowledge and are perfect competitors and this means that lower priced uses are lower *value* uses. We take it as a reasonable assumption of the more complicated reality.

Generally the assumption will be reasonable, but it will impart a downward bias on estimates of the value of open data. The increased search facilitated by negligible transactions costs will probably facilitate the serendipitous discovery of some unanticipated high-value uses.

---

[34] Other information in the DECA (2010) report indicates that the cost of distributing PSI had been higher. The paper reports the costs of the agreement over 5 years to move to open data were EUR 2 million (i.e. EUR 0.4 million per year). Using this latter figure gives a benefit to cost-of-provision ratio of 35 to 1. However, this is unlikely to be indicative of the average result as high value opportunities are more likely to be enacted, studied and reported.

And the value of both existing and new uses will probably be magnified by the strengthened network externalities associated with burgeoning re-use.

We also expect that the average value added lost by transaction costs will be related to the size of these costs. That is, the greater the transaction costs, the greater the average value added that is lost. To aid calculation, we assume that the value added per new re-use under open data is in direct proportion to the size of the transaction costs that are removed in moving to open data.

Using a model described in the appendix, we can estimate the change in GVA as follows

$$\frac{\text{GVA under open-access regime}}{\text{GVA under cost-recovery regime}} = 1 + \frac{e_f + te_o}{2 + 1/e_f}$$

Where:

$t = \frac{p_t}{p_c}$	the ratio of transaction costs to the monetary costs paid by direct users under incremental cost recovery
$e_f = \frac{q_f}{q_c}$	the increase in demand from a cost-recovery regime to a free-but-restricted regime
$e_o = \frac{q_o}{q_c}$	the additional increase in demand from a free-but-restricted regime to an open-access regime

Following's Pollock (2011)'s work, we use an estimate of $e_f = 2$ and, therefore, the above equation can be simplified to:

$$\frac{\text{GVA under open-access regime}}{\text{GVA under cost-recovery regime}} = 1.8 + 0.4te_o$$

Based on other case studies (see section 7.3), we think it reasonable to suggest that e_o is around 2 to 4 (with a midpoint of 3).

For the parameter t, we have found no existing estimates. Based on our experience on similar issues and our own experience in acquiring data, we think it conservative to suggest that these transactions costs are around one-third of the financial costs of a purchase.[35] In such cases, using the above formula, we have a GVA under open data of around 2.2 times the GVA under a cost-recovery regime.

[35] In considering this issue, we considered the time taken to review agreements and the 'mental transaction costs' of adhering to the agreements. In our experience these costs increase with the financial value of the contract and therefore the (average) value of the t parameter may not vary significantly with higher-cost data sets.

If we were to use recent estimates of GVA under a cost-recovery regime of around 0.4 per cent of GDP, then the GVA under an open-access regime would be in the order of 0.9 per cent; that is, an additional 0.5 per cent of GDP.

4.3.1.3 Shifting between other pricing regimes

The above analysis considered the implications of changing from cost recovery to open data. The impact would probably be less moving from open data to cost-recovery pricing because many benefits of open data arise from the way it facilitates the search for new data applications. Once established, many will likely remain.[36]

It is also of interest to consider what might occur when shifting between profit maximisation and open data. As discussed above, profit maximisation may involve more complex pricing strategy – for instance, differentiated pricing such as 'freemium' to encourage re-use amongst lower value users.

Ultimately, the impact of profit maximisation depends on the strategy adopted by the organisation. At one extreme, an organisation introduces a simple charging mechanism that aims to maximise the short-term revenue from the data. At the other extreme, an organisation adopts a strategy that attempts to optimise profits over the longer term and/or across a broader business base.[37] A third possibility is something in between, whereby attempts to implement differentiated pricing result in waste.

Clearly, the change in GVA between these two extremes is large.

---

[36] Note, however, as the example of Amazon book titles in print suggests, that those with an interest can still leave 'money on the table' where transactions costs offset its value sufficiently. Nevertheless, once data has found its way into useful applications, makeshifts will often be found to maintain these arrangements generally by way of renegotiations of access to the data.

[37] For example, Google Maps offers differentiated pricing regimes which encourage re-use by small users and attempts to recover costs from greater abundance of use.

5 Further matters

5.1 Implications across the value chain

How will the value of free and open data accrue through the data value chain introduced in sub-section 2.2?

5.1.1.1 Data publishers

Data publishers could use their market power to maximise profits, increasing profits in the short term. However, we doubt this would be substantial for several reasons. First, the public good nature of digitised data makes it difficult for any publisher to capture much of the consumer surplus generated. To prevent downstream competition between direct customers receiving their data, a data publisher seeks to control distribution, removing competition in downstream markets; however, this would likely lead to problems. If the data provider is not vertically integrated with the data developer, there is a risk of double marginalisation whereby both monopolists attempt to maximise profit and, in combination, reduce the value they obtain.[38] The data provider may attempt to solve the problem by vertical integration; however, this is likely to be relatively inefficient as it results in the data provider undertaking services outside its core capability.

Second, often some form of substitute can be generated. For much public data, there are potentially other (though sometimes more expensive or less efficacious) ways of obtaining substitute data. Thus, for instance, if it is no longer possible to obtain data from traffic authorities on the speed of traffic, or if it has risen in price, one can seek it from mobile phone carriers who can measure the speed of mobile phone movement on the road. Third, there are substantial costs associated with employing charging mechanisms to counter the issues above. The combination of these factors suggests that charging may significantly increase costs whilst reducing demand.

5.1.1.2 Intermediaries

Data aggregators

Data aggregation involves compiling existing open data sources into more useable forms.[39] Typically, aggregators provide basic access for free and charge for higher value-added services. Aggregation appears

---

38 This problem is known as double marginalisation.

39 There are several companies that offer data aggregation services. An informal review of some companies is available at http://www.eveahearn.com/judging-open-data-aggregators/ (accessed 22/1/2016).

to be competitive. There appear to be few material barriers to entry into the market, though we expect substantial investment is required in systems and marketing and, therefore, the primary market will be contested by a discrete number of larger firms with smaller organisations competing in niches.

Given this, we expect a shift to charging would see aggregators negotiating with publishers over pricing in the short term with aggregators' profits falling somewhat. Over time, the market will adjust (e.g. with some aggregators exiting or new entry falling) such that the average profitability of aggregators remains relatively stable.

Enablers

Core data assets are often in a format that developers find difficult to work with. So-called 'enablers' address this problem by further processing the data, for instance, by providing an application program interface (API), a set of routines, protocols, and tools for building software applications.[40] The market for enablers appears similar to that of aggregators, with substantial fixed costs but no barriers to entry and plenty of room for competition and for self-provision amongst its customers.

5.1.1.3 Product and service providers

Broadly, there are two types of product and service providers:

- Developers who create applications for individual consumption
- Data users who use data to enhance existing offerings

In the 'developer' market, there are no material barriers to entry. However, the success of any product may be highly uncertain. The 'data user' category consists of providers of established products. In these markets, the suppliers of products may have some market power. The enhancement to the established products may result in a greater return to those established providers. The 'data user' beneficiaries will typically include other government organisations,[41] who would, we expect, pass on the value to the public through improved services or reduced costs.

For example, the increased re-use of core data assets has increased the value to final consumers from owning smart mobile devices to the benefit (greater producer surplus) of those suppliers of such devices. Nevertheless, competition (or the threat of competition) will typically limit the extent to which such providers will be able to capture the

---

40 An example of an enabler is http://www.transportapi.com/.

41 For example, DECA (2010, pp. 2, 5) concluded that around 30 per cent of the benefits from open access to Danish address data accrued to the public sector.

value. Furthermore, the greater re-use of data will reduce profits of suppliers of products that are displaced by the data re-use. A simple example is that of providers of maps whose business has been transformed by digitisation of data.

5.1.1.4 Consumers

As indicated in this sub-section 5.1.1, competition through the value chain will deliver most of the additional value created from open access to end data consumers.

5.2 Timing

5.2.1 Timing of effects

We outline below a number of what seem reasonable scenarios regarding the timing of market effects. The time period over which the demand impact modelled in section 4.3 should be regarded as reasonably short (in the order of one or two years), suggesting that the full effects of shifting from paid access to open data are felt reasonably quickly.

However, there are some other considerations. The studies referred to changes in direct demand which will include intermediaries and developers. There will be a lag — which may be quite significant — from the time that developers acquire the data to the time that value is realised in the form of products and services and widely adopted in society .

The speed of change will also depend on the direction of change. The studies examined looked at the impact of price *reductions*; none examined the impact of price increases or the introduction of more onerous pricing regimes. The first-round impact of price rises is likely to be faster as the search for value adding uses of the data has already been done. Here the market will move fairly quickly to new price configurations, with some further adjustment as buyers and sellers test each other out and react to counter-party responses.

5.2.2 Accelerating the change

How can we accelerate the change in the economic value when transitioning from paid access to open data? Change may be slow for various reasons. De Vries et al.'s case studies (2011) highlight the importance of removing barriers to reform including reliance on data revenues, organisational constraints and perceived risks to change. They note that public sector bodies relying on PSI sales revenues and value adding appear deadlocked "when there is no other sustainable alternative income stream available".

They also noted, “Further barriers to change relate to statutory provisions imposing cost-recovery schemes, the legacy of old re-use regimes, and the sheer difficulty of changing existing practices”, and noted incumbent re-users with considerable interest in preservation of status quo may try to prevent PSBs lowering charges.

They noted that change could be driven by a top-down process (e.g. by political mandate) or by a bottom-up process (i.e. from within the organisation). In the case of the latter, additional effort was required to justify the reform and secure funding for the transition. Regardless, the study noted “the PSBs interviewed declared that a clear path to transition and the financial means to do so have been of crucial importance”.

And with the market changing fast, measures to deepen market development will also help accelerate the achievement of beneficial results as set out in the following sub-section.

5.2.3 Market development

5.2.3.1 Fostering additional investment in data curation

Some data, such as meteorological and geographical data, is created for its use to those downstream, and so in this sense, is created essentially for its value to users (even if it is rarely created *by* those users). However, other data is often a by-product of other activities – for instance, registration and tax data. In these cases, data may be published without much regard to its usefulness. As a result, those creating and curating the data will have little incentive, and often little knowledge of what uses the data may be best put to, or how further investment in the curation and documentation of the data may add value to downstream users. This data curation will be a public good to those downstream who may use and add value to the data. Accordingly, they should have some role in the governance of data curation and dissemination.

Mechanisms might be developed to allow downstream users to identify and build on opportunities. For example, by PSIHs facilitating feedback mechanisms and performing tasks (at cost) for those prepared to curate and prepare additional data.[42] Additional incentives might be provided by enabling PSIHs to obtain additional funding for further data curation.[43]

---

42 Subject to any privacy, security or other technical concerns, they allow outsiders into their systems to work on the data themselves.

43 For example, by providing subsidies to PSIHs to assist with some of the costs of further data curation by outsiders on the grounds that the resulting benefits cannot be captured entirely by those doing the work. Other opportunities might involve granting PSIH’s time-limited monopoly privileges over the improved

5.2.3.2 Building value

The great data projects driven by the private sector have tended to accumulate around digital artefacts – generally platforms – that generate value that draws in users. Those users then contribute their data.[44] While government agencies do not generally, and should not, seek a competitive advantage over anyone, they should seek to generate value where they can. And those in government in an incumbent position frequently pay too little attention to serving users and generating value as an integral part of their operating strategy. Thus, in addition to making data available, government agencies could give some thought to fostering value creation with that data.

In addition to further investments in curation (discussed in the previous sub-section) governments can seed the development of communities of practice – rather like the community around an open source software project – with an increasing user base generating positive network externalities for all members of the growing community to enjoy. In addition, they may be able to seed projects and/or the development of platforms which might grow into 'data traps'. While this may not sit easily within departments of state with fundamental line responsibilities, certainly more independent agencies tasked with market development like the ODI might seek to pursue such goals possibly in collaboration with other private and public interests. And such initiatives might also sit well with innovation units within line agencies.

It should be noted that such an approach swings the government into the business of using its own assets to *seed* deeper data markets – built not just on PSI and government resources in establishing a platform, but also on private data.[45]

Governments can do this using their own resources to seed platforms, they can help those platforms succeed by nudging or compelling their own agencies to contribute.[46]

---

data and/or given PSIH's capacity to levy stakeholders (e.g. In Australia farmers have statutory powers to collectively levy themselves in Australia to fund public good research).

44 As Matt Turck (http://mattturck.com/2016/01/04/the-power-of-data-network-effects/) has put it, "An approach I particularly like is building a 'data trap'. The idea is to build something that delivers real, tangible value to users from the beginning, and incite them to start contributing their data'.

45 An example is discussed in Gruen (2015), "Innovation? How about TripAdvisor for the arts?" 27th Dec, 2015 in The Age, http://goo.gl/iTjwOj.

46 For example Lateral Economics report commissioned by Omidyar Network, by contributing to the development of standards which draw out others' data because the standard has now enhanced its informational value.

6 References

ACIL Tasman, 2008. "The Value of Spatial Information". Spatial Information Systems Limited. Available at http://www.crcsi.com.au/assets/Resources/7d60411d-0ab9-45be-8d48-ef8dab5abd4a.pdf

Anderson, C., 2009. "Free: The future of a radical price". Random House. Summary available at https://summaries.com/index/Free.pdf.

Brin, David, 2015. "Stop Using Adam Smith and F.A. Hayek to Support Your Political Ideology: The irony of faith in blind markets". Evonomics, 15th December, 2015. At http://evonomics.com/stop-using-adam-smith-and-hayek-to-support/

De Vries, M., Kapff, L., Negreiro Achiaga, M., Wauters, P., Osimo, D., Foley, P., Szkuta, K., O'Connor, J., and Whitehouse, D. "POPSIS – Pricing of Public Sector Information Study." *European Commission Information Society and Media Directorate-General* (2011).

DECA, 2010. "The value of the Danish address data: Social benefits from the 2002 agreement on procuring address data etc. free of charge". Available at http://www.adresse-info.dk/Portals/2/Benefit/Value_Assessment_Danish_Address_Data_UK_2010-07-07b.pdf .

Gruen, Nicholas, 2011. "The David Solomon Lecture: Government 2.0 a couple of years on". Available at http://goo.gl/5qAlsb.

Gruen, Nicholas, 2012. "Telcos reciprocate and market is a net winner". The Sydney Morning Herald, Nov. 14, 2012. At http://www.smh.com.au/business/telcos-reciprocate-and-market-is-a-net-winner-20121113-29adq.html#ixzz41dV7N2eG

Gruen, Nicholas, 2015. "Government as Impresario: Emergent public goods and public private partnerships 2.0". NESTA. At http://www.nesta.org.uk/sites/default/files/government_as_impresario.pdf

Hicks, J. R., 1939. "Value and Capital". Oxford University Press, Oxford.

Houghton, J. W., 2011. "Costs and benefits of data provision". Australian National Data Service.

Lateral Economics, 2014. *Open for Business: How Open Data Can Help Achieve the G20 Growth Target*. Report commissioned by Omidyar Network. June 2014.

Morando, F., Lemma, R., & Basso, S., 2013. "Is there such a thing as free government data"? INTERNET POLICY REVIEW.

McKinsey Global Institute, 2013. "Open data: Unlocking innovation and performance with liquid information". New York.

Odlyzko, Andrew, 2003. "Privacy, Economics, and Price Discrimination on the Internet". Available at http://www.dtc.umn.edu/~odlyzko/doc/privacy.economics.ps

Office of Fair Trading, 2006. "The commercial use of public information (CUPI)". OFT861. *United Kingdom.*

Otteson, James R., 2002. "Adam Smith's marketplace of life". Cambridge University Press.

Pollock, R., 2008. "The economics of public sector information". University of Cambridge, Faculty of Economics.

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*, *5*, 168-193.

Rosen, Rebecca, 2012. "The Missing 20th Century: How Copyright Protection Makes Books Vanish", *Atlantic Monthly*, March 30th. Available at http://goo.gl/Z2xKi

Shakespeare, S., 2013. "Shakespeare Review: An independent review of public sector information". *London: BIS*.

Shirky, Clay, 2008. "Here Comes Everybody: The Power of Organizing Without Organizations". Penguin Press, New York.

Smith, Adam, 1776. "An Inquiry into the Nature and Causes of the Wealth of Nations". Available at http://oll.libertyfund.org/titles/smith-an-inquiry-into-the-nature-and-causes-of-the-wealth-of-nations-cannan-ed-vol-1#lf0206-01_label_153

Szabo, N., 1999. "Micropayments and mental transaction costs". In *2nd Berlin Internet Economics Workshop*.

Varian, Hal. "Economic Value of Google". At http://cdn.oreillystatic.com/en/assets/1/event/57/The%20Economic%20Impact%20of%20Google%20Presentation.pdf.

Vickery, G., 2011. "Review of recent studies on PSI re-use and related market developments". *Information Economics, Paris*.

Worstall, Tim, 2015. "Even Google Doesn't Understand Google's Value to the Economy". Forbes, Sep 17, at http://goo.gl/7fvVbZ.

7 Appendices

7.1 Further details of studies undertaken

Table 2: Scope of studies examined

Study	Time period	Region	Data source	User benefits	Scenario	Units
PIRA (2000)	1998-2000	EU	PSI	Direct + Indirect	Current	GVA
MEPSIR study (2006)	2004-2006	EU25 + Norway	PSI	Direct	Current	GVA (market size for PSI)
ACIL Tasman (2008)	2006-2007	Australia	Spatial info	Direct + indirect	Current	GVA
Pollock (2008)	Surveyed 1996-2007	UK	PSI	Direct + indirect	Released through opening of data	Welfare
Deloitte (2013)	2011-2012	UK	PSI	Direct + Indirect	Current	GVA
McKinsey (2013)	2013	Global	Open data from 7 sectors	Direct + indirect	Potential	Output by sector
Lateral Economics (2014)	2015 to 2020	G20	All open data	Direct + indirect	Potential	GVA

7.2 Technical appendix

We build on a model by Pollock (2008 & 2011), summarised in **Box 2** below, that examined the welfare change from moving between average-cost and marginal-cost pricing regimes.

Box 2: Pollock's (2008 & 2011) model

Pollock estimates the loss of welfare from an average cost pricing regime to a marginal cost regime as

$$\text{Welfare loss}^{47} = \frac{2}{5}\varepsilon\lambda F$$

where:

- F — The fixed costs producing and maintaining the information
- ε — The elasticity of demand as measured when changing from an average (cost-recovery) price to marginal cost pricing, and
- λ — The demand multiplier.

Using a combination of parameter estimates, Pollock (2011) estimates the welfare loss (which we present in terms of GDP):[48]

- Upper-end estimates of λ = 8; ε = 3.5 give gains of approximately 0.3 to 0.4 per cent of GDP annually.
- Mid-range estimates of λ = 5; ε = 2 give gains of approximately 0.11 to 0.13 per cent of GDP annually.

We modify the Pollock model by introducing a kink to the demand curve and explicitly consider the impact of transaction costs. This is illustrated in Figure 4 below, which (similar to the Pollock model) shows the direct demand for data assets. To reflect transaction costs, the figure presents demand in terms of the perceived price, which includes transaction costs.

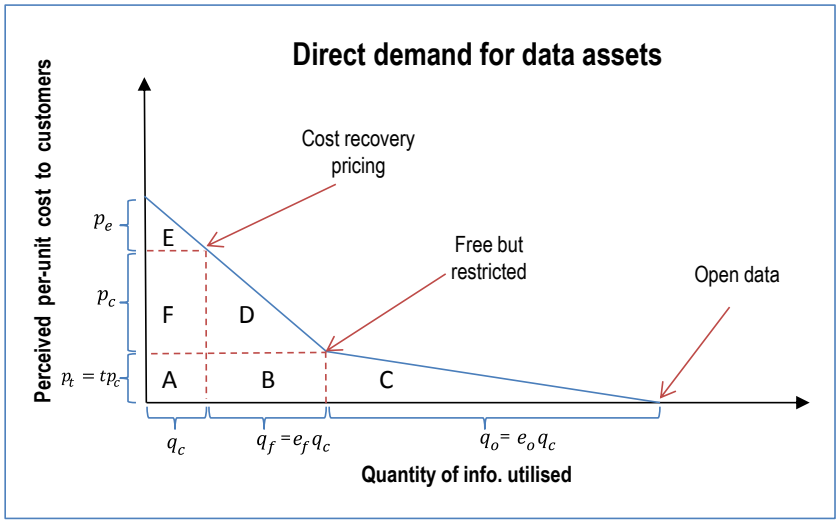
The figure shows a two-part demand curve to better approximate the real demand. While the first part of the demand curve is identical to

47 The '2/5' amount in the equation reflects the assumption that the demand curve is linear and an adjustment for distributional consequences of the subsidy which reduces the welfare loss by a factor of 4/5. Pollock (2008) argues that benefits from lowering the price of PSI are received in proportion to income and, therefore (from a welfare perspective), there is an adverse distributional impact of subsidising PSI. We are sceptical of the need to apply this adjustment. However, in our opinion, an adjustment of similar magnitude is appropriate to account for the marginal excess burden of taxation.

48 These estimates are from Pollock's 2011 paper. These parameter estimates (both for of λ and ε) are higher than suggested in Pollock's earlier (2008) work as the scope of the PSI considered was broader.

Pollock's linear demand curve, the second part is kinked with elasticity rising reflecting the fact that, once the data is open licenced, the distribution of the data becomes 'permissionless', powerfully reducing frictions which would otherwise frustrate the data finding its way to valuable uses.

Figure 4: Demand for data assets



Empirical evidence suggests that the second part of the demand curve is much flatter than the traditional demand curve (which will be discussed in the next subsection). In other words, the transactions costs associated with the chain of permissions to distribute data in a licenced regime deter a sizeable amount of users from utilising data assets. Using this basic approach, we can estimate the proportion of value loss from a paid-access regime using a few parameters:[49]

- the proportionate increase in demand that occurs when moving between pricing regimes, and
- the significance of the transactions costs in proportion to the costs of acquitting the public sector data.

[49] Note: the diagram illustrates a situation in which new users brought into the regime from its move to open licensing will gain relatively low value as they are further down the demand curve. This assumption tends to underestimate the value of increasing demand from permissionless distribution. This is because the transactions costs of licensing frustrate *search* for users and once search costs fall and new uses are found, it seems likely that *some* new uses will turn out to have relatively high value. For instance, some of the users introduced to automated voice directions while driving on Google Map's free system would experience a functionality from the service which, had they known of it before, would have induced them to pay for positively priced services like Navman or Tomtom. This possibility is discounted in our treatment.

In the diagram:

- p_c is the per-user monetary cost of information under cost recovery pricing
- p_t is the per-user transaction cost in a free-but-restricted pricing regime
- q_c is the quantity demanded under cost recovery pricing and q_f and q_o is the additional quantity under free-but-restricted regime and open data.

The additional welfare of moving from cost recovery to marginal cost (free-but-restricted) is the area D times a demand multiplier (λ). If the demand multiplier (λ) is constant across the demand curve, then the GVA of core data asets under cost recovery is λ (E+F) and under a free-but-restricted regime is similarly λ(E+F+D).

In moving to an open-access regime, two additional effects happen. There is an increased demand reflected in the area C. There is also a reduction in transaction costs to existing re-users equal to the area A+B. To determine the increased GVA associated with increased re-use, we have taken a similar approach by multiplying the area C by the same demand multiplier; that is, λC.

The reduction in transaction costs for existing users (area A+B) would have a net-welfare benefit but would not impact on GVA. Similarly, there would be a reduction in transaction costs for suppliers of data assets. Similarly, this reduction would have a net-welfare benefit but no impact on GVA.

The impact of moving from cost recovery to open access is, therefore, to increase GVA from (E+F)λ by the amount (D+C)λ.

As a multiple, the increase in GVA is $1 + \frac{\mathrm{D+C}}{E+F}$

We set up three ratios to help solve the model.

$t = \frac{p_t}{p_c}$	the ratio of transaction costs to monetary costs
$e_f = \frac{q_f}{q_c}$	the increase in demand from a cost-recovery regime to a free-but-restricted regime
$e_o = \frac{q_o}{q_c}$	the increase in demand from a free-but-restricted regime to an open-access regime

The areas C, D, E and F can all be computed as a function of F. These are:

- $C = ½tp_tq_o = ½te_oF$
- $D = ½p_fq_c = ½e_fF$

- $E = \frac{1}{2} p_e q_c = \frac{1}{2} F / e_f$

Therefore, shifting from a cost reflective to an open-licence regime will increase GVA by a multiple of:

$$1 + \frac{\mathrm{D} + \mathrm{C}}{E + F} = 1 + \frac{e_f + t e_o}{2 + 1/e_f}$$

Consistent with Pollock (2011)'s mid-range estimate of elasticity, we assume e_f to be equal to 2 (Of note, his higher estimate is 3.5). As discussed in the next sub-section, we assume that e_o is between 2 and 4 (with a mid-point of 3).

7.3 Evidence of changes in response to price

A number of studies have examined the changes in demand for PSI as a result of changes in prices. Pollock (2008) provides a survey of evidence of price elasticity estimates for PSI. The elasticity analysis in Pollock is complemented by a more recent study — De Vries et al. (2011) — that involved 21 in-depth case studies where public sector bodies (PSB) had changed prices. The case studies were divided into four domains, where the three major domains each encompassed a 100 per cent price-cut case.

A brief summary of the cases reported in these two papers is provided below. While the elasticity estimates in Pollock's papers are within "a large range", the sensitivity of quantity demanded can be alternatively inferred from De Vries et al.'s case studies. Using these results, we developed estimates of the parameters for the modelling. For change in demand from cost-recovery to marginal-cost pricing, we have assumed an increase of 200 per cent. This is consistent with Pollock's (2011) mid-range estimate. To estimate the impact of open data, we more closely examined the changes in demand reported by De Vries et al. (2011). As can be seen from the summaries, there are very large demand increases following price cuts.

When comparing cases, there have been 100 per cent price cuts. With cases with slightly small price cuts, we observe very different changes in the usage increase — with a 100 per cent price cut, the increase is much more significant. In terms of the monetary costs to customers, the 100 per cent price cut cases are similar with (for example) a case where there is a 97 per cent price cut case. Because the latter case is not completely free of charge, there will be transaction costs, which we expect to be the main driver for the usage difference.

In both the Meteorological and the Geographic domain examples provided (from De Vries et al. 2011) below, the *increase* in demand following a 100 per cent price cut was around three times as great. That is, for example, if shifting to close-to-zero prices leads to a 200 per

cent increase, then shifting to free and open access would result in an additional 400 per cent increase (for a total of 200x3 = 600 per cent). The additional increase we observe could also in part be attributed to further price reductions (from near to zero to zero). In conclusion, we think it is reasonable to assume that moving from free-but-restricted to open-access will (in terms of demand) lead to an additional 200 to 400 per cent increase in the demand for core data assets (i.e. suggesting e_o will be between 2 and 4).

In light of the above analysis, we have assumed for a mid-range estimate $e_f = 2$ and $e_o = 3$.

De Vries et al. (2011)

In this study, four domains, Meteorological PSI, Business register PSI, Geographic PSI and Other PSI were examined. A summary of demand change in response to price changes are as follows.

Meteorological PSI:

- KNMI — following an 80 per cent price cut, the number of re-users increased by 1,000 per cent.
- Met.no — following a 100 per cent price cut, the number of re-users grew by 3,000 per cent.

Geographic PSI:

- BEV — following an up-to-97 per cent price cut, usage volume increased, which includes: 250 per cent increase for digital cadastral maps, 200-1,500 per cent increase for cartographic products, 7,000 per cent for digital orthophotos, 250 per cent for the digital elevation model, 1,000 per cent for the digital landscape model, and 100 per cent increase in external-use licenses.
- Spanish Cadastre — following a 100 per cent price cut, the number of digital maps downloads increased by 800 per cent, alphanumeric data downloads increased by 1,900 per cent, total downloads increased by 965 per cent.

Other PSI:

- Destatis — following a 100 per cent price cut, the number of unique visitors increased by 1,800 per cent; and the number of downloads increased by 800 per cent.

From Pollock (2008)

Pollock (2008) documented a number of cases about the sensitivity of demand when there is a price change (see Pollock 2008, for the references).

- The Office of Fair Trading (2006) — estimated an elasticity of 0.3 (lower bound) and 2.2 (upper bound) for New Zealand national mapping data.
- Davies and Slivinski (2005) — estimated an elasticity of 0.3 for demand of weather forecasts. This was considered as a lower bound because it excludes demand coming from intermediaries and the private sector.
- Bedrijvenplatform (2000) — estimated an elasticity of 0.48 (lower bound) and 4.17 (upper bound) for public sector geographic data.
- Making Information Freely Available initiative, Statistics New Zealand — estimated elasticities from lowering prices of 6 for Digital Boundaries Files, 34 for Street Link Files, 1.5 for Small Area Population Estimates.
- The Australian Bureau of Statistics — estimated an elasticity of 2.33 (short-run) and 3.5 (long-run) for ABS statistics.
- The Office of Spatial Data Management in Australia — estimated an elasticity of 1.65 fundamental spatial data.

Pollock also documented evidence from the telecommunication sector, which is to some degree comparable to the information sector, to complement his demand-sensitivity study.

- Hausman et al. (1997) — estimated an elasticity of 1.61 for the introduction of voice messaging and 0.51 for the introduction of mobile phones in the U.S.
- Goolsbee (2006) — estimated an elasticity of 2.75 for broadband in the U.S.
- Kridel et al. (2002) — estimated an elasticity of 1.8 for broadband in the U.S.
- Goolsbee and Klenow (2006) — estimated an elasticity of 1.6 for internet usage.
- Hackl and Westlund (1996) — estimated a range of elasticities from 0.09 to 1.25 for international telecommunications in Sweden.