

THE ECONOMICS OF PUBLIC SECTOR INFORMATION

RUFUS POLLOCK

UNIVERSITY OF CAMBRIDGE

NOVEMBER 26, 2008

ABSTRACT. We give an overview of the economics of ‘public sector information’ (PSI) focusing on the question of funding and regulatory structure. That is: who should pay to maintain public sector information and what regulatory structure should be put in place to support this.

1. INTRODUCTION

This paper examines the economics of ‘Public Sector Information’ (PSI). Public sector information is information held by a public sector organisation, for example a government department or, more generally, any entity which is majority owned and/or controlled by government.¹ To have a convenient term we label the entity holding or providing the information the ‘Public Sector Information Holder’ (PSIH).² Classic examples, of public sector information in most countries would include, among many others:³ geospatial/cadastral data, meteorological information and official statistics. As should be clear from this brief list, while public sector *information* could, in theory, include any piece of ‘information’ produced or held within the public sector it is usually taken only to encompass relatively large and

Emmanuel College and Faculty of Economics, University of Cambridge. Email: rp240@cam.ac.uk or rufus@rufuspollock.org. This paper is licensed under Creative Commons attribution (by) license v3.0 (all jurisdictions). Some portions are based on Pollock et al. (2008) which was commissioned by HM Treasury and BERR and authored in collaboration with my colleagues David Newbery and Lionel Bently.

¹In the UK, for example, several of the major providers of PSI are ‘Trading Funds’. These have a quasi-autonomous position but are 100% owned by government and have a ‘parent’ government department.

²It is perfectly possible, in fact frequently the case, that the holder of information does *not* make it available. It is for this reason that we foreground the ‘holding’ aspect over the ‘provision’ aspect in our terminology. It is also possible that one entity may hold the data while the other makes it available. In this case take the PSIH as denoting the combined entities.

³See Table 1 for a more substantial list.

coherent sets of data and exclude the contents of every memo, pamphlet or web-page the public sector happens to produce.

With the development of the ‘knowledge’ economy, driven largely by advances of digital technology, data plays an increasingly prominent role within our societies, commercially and otherwise. Large and growing businesses have been built on collecting, organizing, analyzing and transacting in data.⁴ Furthermore, almost all businesses, especially those in the services sector, increasingly utilise, and require, a wide variety of data sources to conduct their activities. At the same time, citizens and others have come to depend on, and indeed expect, access to a wide-range of information – be it for planning journeys, tracking changes in their local area (building especially), or keeping up to date with the decisions of their elected representatives.

While much data is supplied from outside the public sector, compared to many other areas of the economy, the public sector plays an unusually prominent role. In many key areas, a public sector organization may be the only, or one among very few, sources of the particular information it provides (e.g. for geospatial and meteorological information). As such the policies adopted regarding maintenance, access and re-use of PSI can have a very significant impact on the economy and society more widely.⁵

The potential importance of (public sector) information can also be gauged from a simple but significant analogy: just as the supply of basic physical infrastructure – power, transport, telecommunications – is essential to the traditional economy, so the supply of basic information ‘infrastructure’ – core datasets in the major areas of geography, weather, transport etc – is essential to the ‘information’ economy. Not only does this comparison provide an indicator of the likely importance of public sector information but it is also illuminating in other ways.

⁴Search engines, today among the most well-known and most profitable enterprises on the planet, would fit squarely within this category.

⁵Quite apart from the immediate competition issues raised by the existence of a government controlled (and often government-mandated) monopoly.

First, core information providers and existing utilities often have similar cost structures where large fixed costs are combined with low marginal costs. Relatedly, many utilities, at least in some areas of their activities, have ‘natural’ monopolies just as PSIHs may do in some areas of their business. Second, utilities are usually providing ‘essential’ infrastructure which, if not directly essential to government, are essential to the general economy – this could be seen as similar to the ‘public task’ of many PSIHs. Third, precisely because of the factors just mentioned, many utilities are regulated and have been for some time. It seems likely that these regulatory experiences can provide useful analogies when considering the situation of PSIHs (few, if any, of which have any independent regulation at the present time).

Even from this brief introduction, it should be evident that the operation of public sector information provision raises a variety of questions – empirical and theoretical; social and economic; regulatory and otherwise. Here, we address many, though not all, of them. We begin with a basic overview of public sector information, what it is, its salient features from an economic point of view, and some important terminology. The next sections then focus on the central issues of funding and regulatory structure. That is: who should pay to maintain PSI and what regulatory structure should be put in place to support this. In particular, we ask which of the three possible groups – users, updaters and government – should bear the burden of paying for the production and maintenance of public sector information. Specifically, is better to (a) make public sector information available at marginal cost (\approx zero) to users and pay for its production and maintenance out of general government revenues or by charging those who to register and/or update it; or (b) fund production and maintenance largely out of charges on those who access and use the information with government contributing no more than is warranted by their role as users.

1.1. What Public Sector Information Do We Consider? We impose three important restrictions on the types of PSI we consider in this paper. First, we restrict ourselves to *digital* information, that is information which can be made

available in digital form (note this does not mean the data was originally collected in digital form, simply that it can be made available in digital form). This assumption ensures that we are always dealing with material whose marginal cost of production/dissemination may be taken to be zero.

Second, we restrict our attention to the provision of *non-personal* information, that is PSI which either contains no personal information or does so at a level of aggregation and anonymization such that personal (private) information cannot be identified. This excludes datasets such as individual tax-records or health data but does not exclude items such as data on property ownership (traditionally publicly available) or even information on vehicle registration if suitably anonymized. As such, non-personal data still includes the great bulk of (socially and commercially) important information. A non-exhaustive list of the types of material we are considering is provided in Table 1.

1.2. Key Features of PSI and PSIHs. There are a few key facts central to any analysis of the maintenance and provision of public sector information. These are, in no particular order: the nonrivalrous nature of public sector information, its associated cost structure (high fixed costs, very low marginal costs), its high potential for use and reuse, and, lastly, the two-sided nature of those who hold and maintain the information. We discuss each of these in turn.

1.2.1. Nonrivalry (Zero Marginal Cost). one person's use of a piece of information does not exclude another from doing so. This equates to the fact that it is (approximately) costless to reproduce a piece of (digital) information once the first 'copy' is made. This contrasts with 'normal' physical goods: if you are using my car I cannot also use it at the same time. However, if one shares a piece of information another gains without any corresponding loss to oneself. Formally, we can also state this as the good having (approximately) zero marginal cost of production.⁶

⁶Production includes copying the data and distributing it to a new user. At the present time, both may have some cost. However, even for large datasets the cost of temporary storage and

Type	PSIH	Comments
Company Information	Companies House	Company registrations, returns etc.
Vehicle Registration	DVLA	Statistical summaries suitably anonymized.
Physical Property	HM Land Registry	Ownership, boundaries, charges etc.
Intellectual ‘Property’	IP Office	Patents, Trademarks etc.
Meteorological Data	Meteorological Office	All forms of weather and climate related information.
Geospatial Information	Ordnance Survey	Traditional ‘mapping’ data but also route and aerial information.
Hydrographic Information	Hydrographic Office	Marine charts etc.
Socioeconomic Statistics	Statistics Authority	GDP, Unemployment, Population etc.
Environmental Data	Environment Agency	Widely varying but including standard pollution and ecological data.
Official Gazettes	OPSI and others	Official notices etc.
Transport statistics	Department of Transport	Journey and planning statistics, public transport information etc

TABLE 1. Examples of Public Sector Information and their Providers in the UK

1.2.2. *High fixed costs.* Collecting, processing and storing data often have substantial fixed costs. Fixed costs are in some ways the ‘flip-side’ of the nonrivalry of information goods: while they cost nothing to reproduce once you have the first copy producing that first copy may be expensive.

1.2.3. *High potential for use and reuse.* Public sector information, like much other information, has the important feature that there are many heterogeneous ways in which it can be used and reused. This potential, especially for reuse in other products and services, relates to the ease with which information can be copied and modified, and it is also a major factor distinguishing it from other goods. For

bandwidth is likely to be very small – and certainly tiny relative to any other cost involving in managing that data.

example, once a piece of steel has been used to make a car there is no easy way for it to be reused elsewhere. However, using a piece of geodata in one particular way in one application does not prevent it being used in a very different way elsewhere. Furthermore if modifications made to the data for one particular use prove valuable elsewhere those changes can be easily, and almost costlessly, shared.

1.2.4. *Two-Sided Nature of PSIHs.* Any information holder can be seen as having two-sides to their operation: the input (write/update) and the output (read/use) side. For example, a registrar of companies must collect the data for its register (input/write) and then may supply this information to third parties (output/read). Similarly, a manager of geospatial information makes changes to their database in response to surveys and changes in the environment (input) and then supply this data to third parties (output). This fact – that all datasets involve both read and write operations – has important implications for policy as it means that: *charges can be made on both sides.*

That is, the revenue needed to create, update and maintain datasets can be levied (in most cases) on *both* the read and write side of the Holder’s operations. That is, *both* those seeking to write (for example register a company) and those seeking to read (get a copy of the dataset or some portion thereof) can be charged. Thus, a policy-maker seeking to fund the production and maintenance of a dataset (or datasets) has three possible options (not mutually exclusive):

- (1) Government funding: fund from general government revenues
- (2) Updater funding: charge those who make changes to the dataset(s)
- (3) User funding: charge those who use the dataset(s)

Which of these should/can be used will depend on the social, technological and political circumstances. In particular, option 2 is not always possible because there are no ‘updaters’ to charge – as with meteorological data for example. For this reason, when discussing funding options we will focus on comparing the first and third options – government funding versus user funding – as these methods are

always available. However, in many cases, option 2 is also feasible. This is important because *politically* it may be easier to alter the balance between 'read' versus 'write' funding than to move to direct payments from government (central or local), particularly if the immediate costs would be significant. We therefore return to a discussion of this particular option at the end of 3.4.

1.3. Upstream and Downstream. One of the primary qualities of information is that the same item can be presented, and reused, in a variety of different forms. For example, the same piece of geodata (perhaps describing the roads and features in a particular neighbourhood) may be available both as part of a large comprehensive dataset or in a printed or digital map. Importantly, while many people, once given access to the basic data, may be able to produce the maps there may be only one source for the original, basic, data (the PSIH). As this distinction will be very important for regulating pricing it is valuable to formalize it. Thus, we say a particular dataset held by a PSIH is:⁷

Upstream: if it cannot be substituted directly from other sources.

Downstream: if it could be provided by another organisation should that organization have access to the relevant upstream information.

Thus, downstream information supplied by a PSIH can be seen as being, at least potentially, in competition with information from other suppliers. By contrast, for upstream information the PSIH is the sole source and faces no significant competition in its supply.⁸ For this reason we largely focus on upstream public sector

⁷This definition is closely related to the OFT's definition of 'unrefined' and 'refined' data in Office of Fair Trading (2006, p.5, para 1.5). In fact, in meaning the two sets of terms are essentially identical. However, we prefer the 'upstream/downstream' distinction for several reasons. First, this is more usual terminology within the competition literature. Second, 'unrefined/refined' has some unfortunate connotations. Specifically 'refining' has obvious suggestions of 'processing' or 'distilling'. But for PSIH data, while, in general, one would expect 'unrefined' data to be fairly 'unprocessed' this need not necessarily be so (after all, almost all data has been processed to some degree to get it into a usable form). 'Upstream/downstream' terminology does not suffer from this defect yet still preserves the underlying meaning.

⁸Note that it is possible that the PSIH is still the sole supplier of downstream information – as long as it *would* be possible for another organization to supply that information it should be classified as downstream. This highlights the importance of having a 'level playing field', in particular any

information on the assumption that a) downstream depends on upstream b) if upstream is correctly managed downstream will, thanks to competition, take care of itself.⁹

1.4. Funding: Charging and Usage Policies. Section 1.2.4 discussed the three major sources of funding for a PSIH: government, updaters and users. These options naturally translate into charging policies – that is prices charged to (external) users and updaters. There are three basic data charging policies for a policy-maker to choose from:

Profit-maximizing: setting prices to maximize profit given the demand faced by the PSIH.¹⁰ Where the product being supplied does not face competition then this will naturally result in monopoly pricing. Here, the usual assumption is that the PSIH will be (more than) fully funded from its revenues and so will require any direct¹¹ government funding – in fact, as a public sector organization, the PSIH will be returning any profits it makes to the government.

Average-cost or Cost-recovery: setting prices equal to average long-run costs (including, for example, all fixed costs related to data production).¹² As with profit-maximization, under this approach it is assumed that the PSIH will not require direct government funding.

other organization should have access to upstream information on the same terms as the PSIH itself and there should be no cross-subsidy between upstream and downstream in the PSIH's operations.

⁹Competition is not, of course, always a panacea. However, it is clear that upstream must be first area to be addressed. If, even after upstream has been addressed, there remain downstream issues these can then be addressed in their turn.

¹⁰One also occasionally hears reference to 'market-based pricing'. It is not entirely clear what this means since several of these pricing strategies involve attention to the structure of the demand curve (that is the price/demand trade-off displayed by the market). However our interpretation is that is intended to indicate that the trading-fund behaves as any other 'normal' market participant would and sets a price to maximize profits given the underlying demand curve.

¹¹Note, however, that as the government may well be a (large) customer of the PSIH this does not mean the government does not make payments to the PSIH – just that these payments are made like 'any' other customer (see Section 2 on Regulation for more discussion of this issue).

¹²There are a various subtleties as to what exactly cost-recovery entails which are discussed further below. See the discussion in the corresponding section of Pollock et al. (2008).

Marginal-cost (Zero-cost): setting prices equal to the short-run marginal cost, that is the cost of supplying data to an extra user. Note that, as we are considering digital data, this cost is essentially zero and marginal cost and zero cost pricing are identical. In this case the PSIHs revenues from maintaining and supplying information will fall below its costs and the PSIH will depend on direct government funding (a ‘subsidy’) to continue its information operations.

When considering the supply of information, price is not the only consideration: in addition one must specify what those who acquire the information can do, in particular, what restrictions there are on reuse and redistribution. In general, and for obvious reasons,¹³ profit-maximizing and cost-recovery pricing are associated with the PSIH retaining strong control over reuse and redistribution – in particular the ability to impose any conditions on reuse and redistribution of its data permitted by the underlying intellectual property rights existing in that material.

For marginal cost pricing, by contrast, it would be natural for the PSIH to make the data ‘openly’ available so that anyone who acquired data would be free to reuse or redistribute in any way they saw fit.¹⁴

¹³If free redistribution and/or reuse any user would immediately be able to compete with the original supplier thereby undermining their ability to charge. It is this logic that lies behind the grant of most IP rights.

¹⁴Note that this would not exclude the imposition of conditions entirely. For example, the PSIH might wish to impose ‘integrity’ conditions so that where its data was supplied by others it was clearly marked as only coming indirectly from the original source and therefore potentially no longer having the same reliability (such a provision already exists with the PSI ‘click-use’ license in the UK). PSIHs might also wish to make certain ‘public-interest’ restrictions. For example, the Land Registry in the UK already prohibits usage of its data for unsolicited mail-shots. Going even further PSIHs could utilize share-alike type licenses of the sort popular in open-source and open-knowledge communities. Here, material is made freely available for use, reuse and redistribution but with the ‘share-alike’ proviso that any derivative work is distributed under the same ‘open’ terms as the original material.

2. REGULATION: COMMITMENT, INCENTIVES AND EFFICIENCY

2.1. Introduction. This section considers how best to address the major regulatory questions raised by public sector information, particularly those related to commitment, incentives and enforcement. As discussed in the main Introduction, there are close analogies between PSIHs (especially the major ones) and traditional utilities, particularly in regard of their cost structures and the role they play in their particular sectors of the economy, and these analogies, and the existing regulatory experience in other areas, will necessarily inform our discussion here.

At the same time, we should note some important differences. Most significant is that, compared with many other ‘regulated’ industries, government takes multiple roles in relation to PSIHs. In particular, government often acts as shareholder, regulator/parent, and customer. Furthermore government’s customer role is far more prominent in relation to PSIHs than in relation to any other ‘utility’ – government is sometimes by far the largest customer for PSIH data and in some cases may account for over 50% of ‘sales’.¹⁵

This close relationship is reflected in the status of PSIHs which are either fully inside ‘government’ or, even when quasi-autonomous, have no separate legal identity. This means, for example, that while PSIHs can draft detailed ‘Memorandums of Understanding’ or ‘Customer Supplier Agreements’ with (other parts of) government it is not clear whether these are legally enforceable contracts – after all, it is not possible for a government to sue itself. This problem is made worse by the fact that the government-to-PSIH relationship is frequently rather opaque, with it being unclear what a given PSIH can and cannot do in relation to product supply, charging etc.

¹⁵For example, in the UK the Met Office income from government whether via sales or subsidy is over 80% of revenue. Even for Ordnance Survey where the proportion of revenue coming from government has been falling the proportion is close to 50% and for particular product ranges may be well over that. At the same time, for some other PSIHs, especially those which are registration-based, the proportion of income from government is very low (approximately zero in the case of the Land Registry, for example).

To give a concrete example, one of the advantages often cited of average-cost or profit-maximizing charging policies for a PSIH is the greater freedom and certainty it gives PSIHs because they need no longer be dependent on direct government funding (discussed further below). However, in many cases (other parts of) government are the major purchaser of data from a PSIH. In this case it is entirely possible for the government to use its role as a monopsonist to reduce suddenly its payments in lean years (just as the government might choose to reduce a subsidy). Conversely, it is not clear what would necessarily prevent a PSIH using its position as a sole supplier of some data products to raise charges to government very sharply. Obviously, in practice, neither of these outcomes are particularly likely, precisely because of the close connection between PSIHs and government. This connection is clearly very important but is, as yet, largely unformalized in most countries.

Finally, a crucial point to bear in mind is that many of the PSIHs enjoy a near-monopoly on at least some of their data, a monopoly, furthermore, made possible or strengthened by government activity. For example, in the UK, for ‘registration-based’ PSIHs such as Companies House, the Land Registry or the DVLA it is a statutory requirement to deposit data with them. In the case of the Met Office, in addition to the natural monopoly afforded by the high fixed costs of data collection, the government provides substantial funding for the PWS.¹⁶ Furthermore, in most cases the data marketplace in which PSIHs operate have a clear upstream/downstream structure with the PSIH ‘monopoly’ most prominent in the upstream market. This presents a whole raft of competition issues, particularly in relation to tying, exclusionary dealing, predatory pricing and the like. As a result it would seem clear that some form of price/access regulation would be necessary if abuses of market power were to be avoided, and adequate competition and innovation be encouraged (at least downstream). It would also make it extremely difficult to permit PSIHs to

¹⁶The Met Office have sought to address some of the problems these may cause from a competition perspective by maintaining a clear division between their ‘wholesale’ and ‘retail’ arms with the same access terms applied to all, including their own retail division, when purchasing data from the ‘wholesale’ arm.

pursue a profit-maximization (monopoly-pricing) strategy in the absence of some form of regulatory oversight.

2.2. Commitment. In most analysis of funding structures, including that below, it is explicitly assumed that government would provide any necessary subsidy to maintain PSIH income at a level sufficient to maintain the relevant dataset (should a charging policy be chosen that resulted in PSIHs income dropping below costs). This *implicitly* assumes an ability for government to *commit* to payments both now and in the future. Such an ability cannot be taken for granted. Governments around the world have frequently demonstrated the difficulty of making such commitments and the impact of political considerations on infrastructure investment.¹⁷ Sudden fluctuations, or simply reductions, in the level of subsidy would be likely to have substantial negative effects on the ability of PSIHs to maintain both the range and quality of their information. Clearly, the issue of commitment is an important one to consider.

The issue of commitment is not solely confined to the case where subsidies are being provided. Consider, for example, the hypothetical situation where a PSIH is following a policy of profit-maximization but still retains its current institutional setup where it sits within the public sector. Suppose then that the PSIH decides that one obvious way to increase profits is to increase charges to central and local government, perhaps to the extent that some subsections are no longer able to purchase the data. In this case there might be substantial pressure brought to bear by government on the PSIH to price more ‘reasonably’, or government might consider amending the PSIHs charging policy. In either case the government would have reversed its ‘commitment’ to allow the PSIH to pursue a policy of profit-maximization. Thus it should also be clear that while the ‘commitment’ issue may be most prominent in the case where government is providing funds it arises in

¹⁷For example, in 1991, the UK government promised an extra 750 million pounds to the Tube to do renovation work only to have to reverse this commitment a year later due to sudden pressure on the national finances. (LRB, Vol. 27 No. 9, 5 May 2005). See also the discussion of the Land Registry’s experience in the early 1990s below.

relation to all of the possible pricing policies. In fact, as discussed further below, the commitment issue relates more to the institutional and regulatory structure in which PSIHs operate than to the chosen charging policy.¹⁸

2.3. Incentives and Efficiency. In addition to the basic commitment issues it is also the case that different charging policies, and the associated different relationships with central government, might result in different incentives faced by PSIHs. In particular charging policy could affect incentives for responsiveness, innovation (development of new products), cost reduction and general performance.

For example, a PSIH which has been mandated to price data products at marginal cost may have reduced incentives to develop new products as it will not be able to reap any particular benefits from doing so.¹⁹ Conversely, if marginal cost pricing was combined with some kind of per unit output subsidy this could result in incentives for *over-investment* in quality and capacity improvements because, by over-investing, the PSIH stimulates demand and obtains a larger subsidy.

In terms of responsiveness an organization operating a more ‘commercial’ pricing policy (e.g. profit-maximizing) might lead a PSIH to be more customer oriented – more responsive to complaints and more concerned about general service quality.

Similarly, wherever a PSIH is regulated (i.e. in all cases except profit-maximization) it may lack adequate incentives to reduce costs – because any reduction in costs may be partially appropriated by the regulator (either in the form of a lower subsidy or lower prices).

2.4. Information and Enforcement. All of the charging policies considered with the exception of profit-maximization require some form of regulation (by government or otherwise) to ensure compliance. Even in the case of profit-maximization the government’s role as sole shareholder would necessitate some form of oversight.

¹⁸See also the discussion of the government’s multiple roles above.

¹⁹The same could be true in theory from average cost pricing though this depends somewhat on the degree to which the organization engages in cost recovery at the organizational rather than the per product level.

One might assume that marginal cost (and zero-cost) pricing would require more information (and more effort on the regulator's part) than average cost pricing. In particular as it is unlikely that the level of investment is constant over time there will be important questions as to how subsidies (and price regulation) were allowed to change over time to reflect these needs.

However, as already alluded to above, under cost-recovery managers may have an incentive to 'over-invest' since higher costs can be covered by increasing revenues ('gold-plating'). Additionally, with the ability to set prices in at least some areas PSIHs could also behave inefficiently, for example, by investing in poor projects, while still complying with cost-recovery at the organizational level since losses could be made up by raising prices or cross-subsidies from other parts of the business. The information needed by a regulator to avoid these outcomes is similar to that required when monitoring a marginal-cost or zero-cost regime – in particular the regulator will need to monitor investments in order to ensure that they are at the efficient level.

Leaving aside these investment questions it is certainly true that different pricing regimes provide different information about the demand curve (and therefore implicitly about surplus).²⁰ Specifically, if the given pricing policy is being pursued at the per-product level, then profit-maximization and average-cost both have the advantage that they guarantee that a given product is only produced if the surplus from doing so is positive. By contrast under marginal cost pricing it is possible for a product to be produced (and subsidised) whose net surplus is negative. However it should be noted that this particular point can be taken both ways. A profit-maximization or average-cost regime ensures that a product is produced if and only if the producer surplus is positive (i.e. revenues are larger than costs). Thus there may be products whose total (consumer plus producer) surplus is positive – and therefore worth producing – but whose producer surplus is negative. These then are

²⁰This will also be discussed below, see figure 1 in particular.

products which might well be produced under a marginal cost regime but would not be under an average cost or profit-maximizing regime.

2.5. Discussion. The main points of the previous sections are drawn together in Table 2. One important possibility to bear in mind when reading this, and when considering these issues in general, is the likelihood that any given charging rate might be applied selectively. For example, different charging policies could be applied to upstream and downstream data – say, marginal cost for upstream and average cost, or profit maximization, for downstream. Thus rather than situating a PSIH in a single column it is important to keep in mind that it could be ‘spread’ across several, with different parts of a PSIH’s operations under different charging policies. The table attempts to reflect this, at least to some extent, by explicitly noting where a particular point relates only to data with particular properties.

There are two major lessons to take from all of this. First: there is no direct linkage of charging policy to governance issues – in fact governance questions are best seen as orthogonal to pricing ones. In particular, all policies require some form of regulation to function well. Second, and relatedly: charging policy is not the central issue when considering problems such as commitment and incentives which are the primary determinants of performance in terms of data quality, investment and efficiency. Rather, charging policy is best seen as secondary, and dependent upon, the primary matter of the regulatory/governance structures under which data provision (and collection) by PSIHs occurs.

2.5.1. Commitment. To illustrate, consider a concrete example provided by the Land Registry, one of the major UK Trading Funds. In the late 1980s and early 1990s just prior to becoming a Trading Fund, the Land Registry operated a cost-recovery regime in which charges were set to cover costs. However, it did not control its revenues but rather returned them to central government. The Land Registry management would then go ‘cap in hand’ to negotiate their budget for the next financial year. According to them, in the late 1980s this resulted in a degree of

Issue	Profit Max.	Avg. Cost	Marginal/Zero Cost
Commitment	Good. ^a	Largely dependent on regulatory / governance structure.	Largely dependent on regulatory / governance structure. ^b
Incentives	Optimal for PSIH though likely non-optimal for other market participants (see next item).	Risk of over-investment and inefficiency (costs too high). Monitoring required of investment, quality and costs.	Risk of either over or under performance depending on subsidy function. Monitoring required of investment, quality and costs.
Distortion of Competition	Upstream: major issue given dominant position of PSIHs. Downstream: minor as long as cross-subsidy is limited.	Significant issue if PSIH provides internal access to upstream material on different terms to external firms (esp. if cost allocation between upstream and downstream is opaque). ^c	Minor. ^d
Information	Not relevant as no regulation.	Single point on demand curve where revenue covers costs. ^e At aggregate level know PSIH covers total costs.	Single point on demand curve where price equals marginal cost.

^aThough could depend on relationship of government and PSIH – particularly risk that profits are ex-post ‘appropriated’

^bCould be a greater issue than under ‘average-cost’ because here the government may be providing subsidies.

^cOversight would still be required here to prevent the use of discriminatory tariffs. For example, a PSIH could set a tariff consisting of a one-off, but very large, fee for all its data. This might then exclude external users who only need a small part of that data. Similarly without transparent cost allocation under average cost pricing a PSIH might have an incentive to overcharge for upstream access to exclude downstream entrants – a problem familiar from the telecommunications literature, see e.g. Farrell (2003).

^dThough the provision of subsidy may retard entrants who wish to compete directly with the PSIH in the provision of data. However, as long as the marginal cost data provision were largely confined to those datasets of which the PSIH was sole provider this would not become an issue.

^eThough where a PSIH performs cost-recovery only at the aggregate level the exact relation of revenue to costs for a given product may be unclear.

TABLE 2. Charging Policies and Regulatory/Governance Issues.

underfunding, which made it impossible to deal with the level of applications they were receiving. As a result a large backlog of applications built up.

In 1993 they became a Trading Fund, in part because of the problems that had been encountered. Since then this sort of problem has not recurred and, in their opinion, the greater autonomy provided by being a Trading Fund means that investment can be planned better and they are less subject to the vagaries of ‘vote-funding’.²¹ Note that throughout the basic charging policy was unchanged with cost-recovery both before and after Trading Fund status was obtained. Hence, here it would seem clear that if the improvements in service quality were due to anything, they were due to changes in the regulatory environment, in particular the greater certainty and autonomy provided by the Trading Fund structure.

To take this point further, whenever PSIH funding is directly controlled by government, there will be potential commitment issues under all pricing regimes (see discussion above). Moving to a different regulatory structure could improve this. For example, if PSIHs were more legally independent it would permit the creation of arm’s length legally-binding contracts regarding both subsidies and purchases. Combined with independent and transparent regulation this sort of structure would go a long way to eliminating concerns about the ability of government to deliver on subsidy and purchase promises and eliminate fears about the effects of such risks on the quality and availability of PSIH data.

In particular, it should be emphasized that a change in charging policy, for example to use marginal cost pricing for some part of a PSIH’s products, does *not* require removing a PSIH’s independence or a reversion to ‘vote-funding’. In fact, as

²¹Though interestingly all of their fees are still set by government through fees orders (more precisely the fees are set by the Lord Chancellor and then approved by HM Treasury). Thus government still largely controls their year to year revenues (and hence, one would imagine, their investment levels and incentives). This suggests that, in this case, the major benefit of Trading Fund status was not to reduce the level of (central) government control but to reduce the risk that government would, especially in ‘difficult times’, take too great a share of Land Registry revenues for other purposes leaving the Land Registry with insufficient funds to carry on its operations. In this sense Trading Fund status could be seen as a form of ‘ring-fencing’ in relation to the Land Registry budget.

just suggested, such a change would optimally be combined with improvements in the independence and transparency of the governance structures to provide PSIHs (and government) with *more* certainty, clarity and independence than they currently have.²²

2.5.2. *Incentives and Efficiency.* Coming to the question of incentives and performance, the differences between charging regimes are, if anything, even less significant than when considering commitment. As already discussed, without adequate regulatory/governance structures in place, all charging regimes can result in poor incentives, inefficiency and overall poor performance.²³ Conversely with a good regulatory/governance structure in place any of the charging policies could be implemented without jeopardising the incentives, efficiency, and performance of a PSIH.

Consider the current situation in many jurisdictions for some of the larger PSIHs, which roughly approximates to capital-based regulation – a PSIH is expected to cover costs and make some specified return on capital. As is well known, this approach has significant incentive and efficiency problems. First, and most obviously, the organization no longer has incentives to minimize costs but rather seeks to match costs to revenue. Furthermore, given the market power PSIHs have, at least in some markets, overspending can always be addressed by raising prices and increasing revenue.

Second, and relatedly, the organization now seeks to equate average costs and average revenue rather than marginal costs and marginal revenue. As a result there can be ‘gold-plating’ and over-investment in quality.²⁴ Third, and more subtly, this

²²An obvious example in this respect is provided by the case of the Train Operating Companies (TOCs) formed in the UK post-privatization of British Rail. Here the government has been able to agree subsidies as well as payments for long-term investment. While PSIHs obviously differ from the TOCs in several respects, notably by not being privately owned companies, it would not be very difficult to design mechanisms for PSIHs which that could provide a similar degree of certainty.

²³If the PSIH is still government owned profit-maximization here is no different since the monitoring role usually played by shareholders and the market is now the responsibility of government.

²⁴This is distinct from the previous point in that, for any given project, the costs may be at their, optimal, minimal level for the quality chosen, but that quality will be at inefficiently high level. To put this in terms of a simple example, suppose a purchase of a computer system is being considered.

pricing policy provides incentives to over-invest in order to extend (inefficiently) the capital base since this then allows an increase in revenues.

These are all fairly serious issues. Thus, the government, in its role as owner and regulator of a PSIH, needs to exert a substantial degree of effort to try and reduce or eliminate these risks. In particular, to correct these potential biases in a PSIH's behaviour it would likely need both to put in place some form of incentive scheme, and associated monitoring mechanisms. This has been the approach in other areas. For example, Network Rail (which replaced the privately-owned RailTrack in the UK), though run as a not-for-profit company limited by guarantee has put in place a fairly complex incentives package for managers and is also monitored by the Office of Rail Regulation.

A similar approach could also be taken if a marginal cost pricing regime were adopted. Here too there are issues though it will be assumed that a subsidy can be provided in a transparent and committable way (see the extensive discussion above). Specifically, just as with average-cost pricing, the government (or the regulator if independent) would need to think carefully about providing incentives for (efficient) reduction in costs (while keeping investment at the optimal level). To put this in more concrete terms, for those products priced at marginal cost the government (or regulator if distinct from government) would need to be setting a subsidy level. This subsidy would likely be tied to (previous) output and expenditure in some manner. One option would be to set the subsidy to equal fixed costs in the last period. However, this might result in poor incentives to lower costs. Similarly setting a straight per output subsidy might lead to over-investment.

There are two manufacturers M and N and both offer a high and medium quality system. The two manufacturer's systems are equally good but N's one costs more. In addition, M's high quality system just breaks even (revenues equal costs) while the medium quality system results in a profit (revenues exceed costs). Here then, inefficiency in the first sense would be to choose the N system over the M system resulting in simple over-payment. Inefficiency in the second sense would be choose the high quality system – profits are zero in this case but would have been higher with the medium quality system.

With a little effort and by combining these two approaches one could develop something a lot better. To provide just one example; one could estimate a particular period's fixed costs using previous periods' fixed costs (multiplied perhaps by a deflator), and then use this, together with some estimate of the value of usage, to set the per unit of output subsidy. The reason for incorporating output is that this would ensure a PSIH has incentives to get their data used (whether by making it easy to use, publicizing it etc etc). Additionally, by incorporating output measures this approach makes it easier to allow for the introduction of new data products – which is an important factor to consider when managing marginal cost pricing.²⁵

This is just one example, and one that a regulator would clearly need to consider in more detail. However it should be sufficient to demonstrate that the problems are not insurmountable, and are, in many ways, little different from the issues confronting government when it uses a cost-recovery approach.²⁶ What is clear in both cases is that the incentive questions must be addressed. If they are not, there would most likely be serious detrimental impacts on efficiency and general performance. However as long as reasonable thought and effort are put into dealing with these issues, in particular by designing a robust governance/regulatory regime, these negative consequences can be avoided.

2.6. Conclusion. Much of the concern about the impact of a change in charging policy (particularly to marginal cost or zero cost) is based on a misidentification of

²⁵There are other ways to address this. For example one could follow a system used by the TOCs who present a 'shopping-list' to government of possible capital improvement projects which government then chooses from. Alternatively one could provide some way for users to feed back requirements to PSIH regarding new datasets to collect. This is also a major advantage to having a PSIH retain a 'Retail' arm in addition to any marginal cost 'Wholesale' arm as 'Retail' can pass on feedback regarding their requirements to 'Wholesale' (in fact, the UK Met Office stated that something like this already occurs with their 'Retail' division passing back feedback to 'Wholesale' as to what new kinds of data would be useful in the provision of their own products and services).

²⁶Though the very fact that, because of the need for a subsidy, these kind of calculations are more out 'in the open' is a significant advantage of marginal cost pricing. Such an increase in transparency benefits all concerned, and, furthermore, requires that a regulator have access to the relevant cost and output data from a PSIH on a regular basis, and thus, could be seen as a way of credibly committing the government to a more transparent and active governance/regulatory regime in future.

charging policy with regulatory structure. Having a PSIH dependent on year-to-year ‘vote-funding’ for its activities might well have substantial negative impacts – but it would do so whatever charging policy was being followed. Conversely, any of the charging policies discussed could be used successfully if a independent, transparent and coherent governance structure were in place. In this regard charging policy can largely be seen as *orthogonal* to the question of PSIH performance – whether evaluated in terms of quality, responsiveness or efficiency. Moreover, the importance of having an adequate governance structure – whatever charging policy is chosen – cannot be overemphasized.

In many countries some of that structure is already in place. However, as already discussed, there are likely to be several important ways in which it could be extended in pursuit of delivering on the key goals of transparency, certainty, independence and incentivisation. If an adequate structure is in place, and economists and regulators experience over the last few decades provides plentiful experience in this regard, then there is every reason to be confident that almost any pricing policies can be *implemented* without significant adverse effects on the efficiency and performance of the PSIHs affected.

3. WHICH FUNDING MODEL?

3.1. Introduction. When deciding which funding model is ‘best’ we need to know what ‘best’ means. The main ‘outcome’ variables one would consider here are:

Consumer surplus: the value (utility) end consumers derive over and above any payments they make.

Producer surplus: surplus to producers (profits).

Government: Revenue/Expenditure.

Total welfare: sum of these.²⁷

For economists, and most policy-makers, it will be the last of these, total social welfare, which would be the most significant since it is an overall measure which

²⁷Taking account of the cost of government funds.

incorporates all of the other changes into a single value (usually presented in monetary terms for convenience of comprehension). To decide which funding model is optimal requires simply that we perform a standard ‘social cost-benefit analysis’. Conceptually, all this involves is summing up the benefits and costs from each particular option and seeing which one does best. Of course there are some theoretical subtleties, particularly in relation to making adjustments based on who gains the benefits and who bears the costs. However, the major challenge will be an empirical one: obtaining estimates of the major parameters upon which the calculations depend.

One last point before we embark on the formal analysis: in what follows we shall concentrate solely on comparing two of our three charging options: average cost to marginal cost.²⁸ This is not a great restriction for two reasons. First, we are concentrating on upstream material. It would therefore be difficult to allow a PSIH to pursue a profit-maximization regime without raising a host of serious competition issues.²⁹ Second, if marginal cost is superior to average cost charging then it is also (a fortiori) superior to profit-maximization. Hence, the comparison of marginal cost to average cost already yields most of the relevant information we need.³⁰

3.2. Theory. The theoretical underpinnings of the calculations conducted in this report can best be understood by the digram presented in Figure 1. Here, we show the demand curve for a single information ‘product’ which a PSIH could supply.³¹ This (linear) demand function is shown together with the marginal and average cost curves. As illustrated the cost curves correspond to a good having constant marginal costs approximately equal to zero and a non-zero fixed cost of production.

²⁸Those who want a full analysis of all of the charging options can find one in Pollock et al. (2008).

²⁹See the previous section on regulation.

³⁰Of course, if average cost is superior to marginal cost we are left with the possibility that profit-maximization is even better. However: a) there are good theoretical reasons why this is unlikely (deadweight losses grow as the square of the price increase) b) as mentioned, this is problematic from a competition perspective. Furthermore, as we shall see, it is likely that in most cases marginal cost is superior to average cost and so this issue is moot.

³¹Implicitly this kind of partial equilibrium analysis assumes that the prices of other goods (and other information supplied by the PSIH) are being kept fixed while we analyse this particular item.

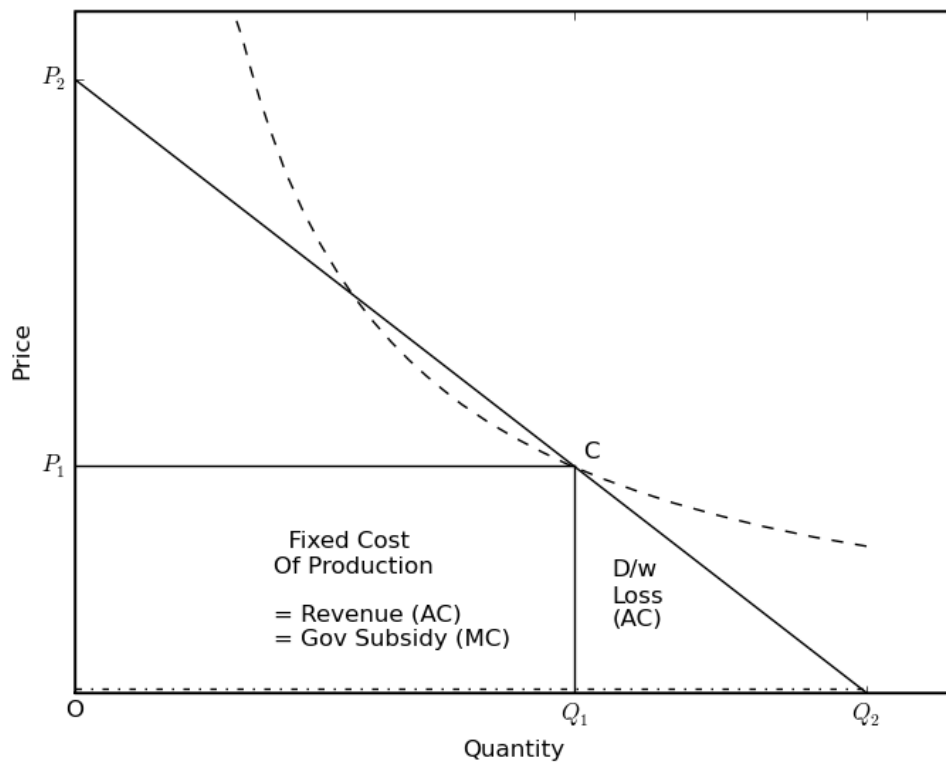


FIGURE 1. Illustrative demand and cost functions. Shown is a linear demand curve for a product with fixed costs and constant marginal costs (approximately equal to 0). Marginal cost (dot-dashed at very bottom of figure) and average cost curves (dashed) are shown.

We would emphasize that the particular functional forms and parameters have been chosen simply for illustrative purposes and do not necessarily indicate those that will be used in doing calculations – though, of course, the natural division of costs into fixed and marginal will be retained.

Table 3 explicitly relates each outcome variable to a particular area under the demand curve in Figure 1. Producer surplus equals profits: that is revenue minus costs (fixed as well as variable). Thus producer surplus is zero under average cost pricing (this is the definition of average cost pricing), and is negative under marginal cost pricing. Consumer surplus, using the partial equilibrium approach adopted here, will equal the area under the demand curve which is above the price being set. Since the PSIH is government controlled, producer surplus equals government

	Average Cost	Marginal Cost
Consumer Surplus	P_1P_2C	OP_2Q_2
Producer Surplus	0	$-OP_1CQ_1$
Deadweight Loss	CQ_1Q_2	0

TABLE 3. Outcomes Under Different Charging Regimes With Reference to Figure 1. Government Revenue has been omitted as it is equal to producer surplus.

revenue/expenditure. Thus under average cost the government receives and expends nothing but under marginal cost must supply a subsidy to cover the fixed costs of producing and maintaining the information.

Finally, to work out social welfare we then sum up consumer and producer surplus. Here, rather than looking at overall welfare, it is simpler to compare the differences between the two options. The situation, at least conceptually, is extremely simple with a move from an average to a marginal cost charging policy involving two changes:

Cost: The government must supply the funds to pay the fixed cost producing and maintaining the information.

Benefit: Users gain surplus equal to this fixed cost plus the deadweight loss.

Now, if funds/surplus in the hands of the government and in the hands of users were equivalent it would be immediately obvious that a marginal cost regime was better – the fixed cost would net out and one would be left with the gain of the deadweight loss. However, things are not so simple: we need to take account of the benefits those government funds would otherwise have generated (if they were not being used for the subsidy). The basic approach for performing this kind of cost/benefit analysis is well known. It involves taking uncommitted government funds as the numeraire and then adjusting the surplus from the project under consideration using the appropriate social weights to reflect the different values of public and private costs and benefits. This surplus is then compared to a standard benchmark project upon which funds could otherwise have been spent (1 unit equally

distributed). In essence this is asking: are the benefits derived from spending government funds in order to have a marginal cost charging policy for this PSI greater than those obtained by spending those government funds on the benchmark project. If the answer is yes the project is worthwhile; if the answer is no the project is not.³²

3.2.1. *The Multiplier.* The approach laid out uses the standard partial equilibrium approach of equating areas under the demand curve with social surplus. But one needs to ask here whether, in this case, demand accurately reflects surplus. Note that this is *not* about the standard question as to whether using the uncompensated (Marshallian) demand curve is a good approximation to the compensated demand curve (see Willig (1976); Hausman (1981)). Rather it is the question whether, for the information goods considered here, the demand curve systematically misrepresents willingness-to-pay and hence welfare. There are two major reasons why the answer to this question is likely to be an affirmative one in the case of public sector information (further discussion of both of these possibilities may be found in the appendix):

- (1) Public sector information is often *not sold directly to consumers but to other firms who in turn provide products (informational or otherwise) to consumers*. As such, the demand curve observed by the PSIH may significantly understate the true value being generated either because downstream firms do not capture the full surplus from their activities or because the downstream market is itself imperfectly competitive.
- (2) The standard demand curve is static, frozen at a particular point in time, with no allowance for how it might change, and, in particular, how reductions

³²An alternative approach involves taking money in consumer's pockets as the numeraire. In that case one needs to determine the marginal cost of public funds (that is, how much does raising one pound of government funds cost general society at the margin).

in *present* prices may, by stimulating the development of products and services both downstream and in other markets,³³ have a major positive impact on *future* surplus.

Both of these two factors provide reasons to think that using the basic demand curve may lead to underestimates of the gains from lower prices – equivalently, underestimates of the deadweight losses of higher prices. This would imply that, when doing cost/benefit style calculations of social welfare, one would need to scale up the welfare related to increases in usage of PSI by some form of ‘multiplier’. We therefore introduce such a ‘multiplier’ parameter into our calculations below.

3.2.2. The Form of the Demand Curve, Regime Change and Transaction Costs. The very limited availability of empirical data, necessitates some assumption about the shape of the demand curve. The approach adopted here will be to assume that, at least in the region of interest, the demand curve may be approximated by a linear function and thus that the elasticity of demand captures sufficient information for us to calculate changes in consumer and producer surplus. For small changes in prices such an approximation is quite reasonable. Of course here the price changes under consideration are likely to be quite substantial. In this case using a more convex inverse demand function (e.g. $p = 1/q$) or a more a concave one (e.g. $p = k - q^2$) might lead to changes in the surplus estimates. Nevertheless, given the data constraints an assumption of linearity seems a reasonable first-order approximation.

We shall also assume that altering charging policy does not change the costs of a PSIH. In reality, it is likely that a change from average to marginal cost pricing would result in a reduction in the costs incurred by the PSIH in creating and maintaining the information – for example, with marginal costs at zero (so the information is provided free) there may be significantly less administrative overhead in relation to billing, contract monitoring, enforcement etc. However, while such cost changes

³³For example, cheaper geodata may lead to more rapid improvement in the quality of the software and hardware components of Geographical Information Systems (GIS). Similarly, Weiss (2004) argues that marginal cost access to weather data in the US was a large factor in the development of the multi-billion dollar weather derivatives industry.

may not be negligible, we will ignore them here for three reasons. First, such cost changes are currently impossible to calculate given the data available. Second, such cost changes are probably ‘second-order’, that is small relative to the main effects. Third, and perhaps most decisively, such an omission is ‘conservative’, in the sense that it biases the results towards the average-cost regime (which is currently is the default in many countries). While inserting ‘bias’ is never first-best, inserting a ‘conservative’ one could be seen here as a reasonable ‘second-best’ – and where a marginal cost (or zero) price cost regime is found to be preferable, this ‘bias’ would be irrelevant in the sense that it would make the preferability ‘stronger’.

Finally, transaction costs will also be ignored, whether these relate to transitioning to a new charging regime or to running a given regime. Our reasons for doing so are similar to those just discussed for general costs. First, there is absolutely no data on which to base estimates of their magnitude. Second, in the case of the costs of transition these are likely to be small, at least compared to the magnitude of the other sums involved. Third, for general transaction costs a move from an average cost to marginal cost regime would likely result in a reduction due to less need for monitoring and enforcement. Thus, ignoring them can either be seen as having little effect or as instilling a ‘conservative’ bias in favour of the existing regime. Again, inserting such ‘bias’ is not first-best, but given its ‘conservative’ nature it could be seen as a reasonable ‘second-best’.

3.2.3. *Algebra.* This subsection converts the preceding discussion into an equation which characterises the welfare difference between the average and marginal cost regimes in terms of the key underlying variables (listed in Table 4). The numeraire for all of these calculations will be government funds (and not funds in the hands of consumers). The choice of numeraire has no effect on the signs of any value and therefore on choice of policy, but simply acts to scale outcome values. Taking government funds as the numeraire seems the natural approach here given their

Name	Variable
θ	Distributional weight for the project under consideration.
$1 + \alpha$	The ‘marginal cost of public funds’. Note that θ and α are linked via $\alpha = \frac{1-\theta}{\theta}$.
F	Revenue under average cost pricing (equal to fixed costs of the PSIH in producing and maintaining the information).
DWL	Size of deadweight loss under average cost pricing.
g	The proportion of revenue derived from government sources under average cost pricing.
λ	Demand curve ‘multiplier’. Note that $\lambda \geq 1$.
ϵ	(Absolute) elasticity of demand at price under an average cost regime.
p, q	Price and output under average cost pricing (point C in figure 1).
Δq	The (absolute) change in quantity (usage) as a result of moving from average cost to marginal cost pricing.

TABLE 4. Key Variables.

centrality in the calculations – it is government funds that will be used in paying any subsidy.

As already discussed at the end of Section 3.1 the change from the average cost to marginal cost regime has two simple effects:³⁴

Cost: The government must supply the fixed cost F

Benefit: Users gain surplus equal to $F + DWL$

The next step is to ensure that both of these terms are represented in terms of the numeraire which is government funds. Obviously government expenditure need not be modified but any gains to those outside government need to be scaled by the distributional weight θ . Thus the first step is to breakdown the costs and benefits into those accruing to the government, and those accruing outside government, whether to consumers (consumer surplus) or to producers (producer surplus). The breakdown is show in Table 3.2.3. The important point is that the values in the un-weighted subsection are not necessarily commensurable since they are not expressed

³⁴We should note here that, in addition to the items already mentioned, we will also ignore issues such as a) some of ‘consumer’ surplus is really accruing to firms and therefore may flow back to the government as tax (and may have a different distributional weight) b) a time differential in the impact of benefits and costs. These are likely to have a relatively small effect and the reader seeking a full treatment can see Pollock et al. (2008).

with respect to the same numeraire. Those in the second ‘Weighted’ subsection have been corrected with the necessary distributional weights to ensure they are all expressed in terms of the numeraire used (government funds).

Item	Expression
Unweighted (no common numeraire)	
Cost:	$-F$
Benefit	$F + DWL$
o/w Govt	gF
o/w Non-Govt Surplus	$(1 - g)F + DWL$
Weighted (numeraire = govt funds)	
Δ Government: ΔG	$-(1 - g)F$
Δ Consumer Surplus: ΔCS	$\theta((1 - g)F + DWL)$
Δ Total Welfare: ΔW	$-(1 - \theta)(1 - g)F + \theta DWL$

TABLE 5. Theoretical Breakdown of Surplus

Our last step is to relate the PSIH fixed cost: F , and the deadweight loss: DWL . The fixed cost equals price times quantity under average cost pricing: $F = pq$. The deadweight loss is more complex and its exact size will depend on the shape of the demand curve. Using the linear form for the demand curve the expression for the deadweight loss takes a particularly simple form as follows:

$$DWL = \lambda \cdot \text{Triangle } CQ_1Q_2 = \lambda \frac{1}{2} p \Delta q = \lambda pq \frac{\Delta q}{2q} = \lambda F \frac{\Delta q}{2q}$$

It will be useful to rewrite this in terms of the elasticity using the fact that $\Delta q/q = \epsilon \Delta p/p$:³⁵

$$DWL = F \lambda \epsilon \frac{\Delta p}{2p}$$

Noting that $\Delta p = p$ (as marginal cost = 0) we have $DWL = F \frac{\lambda \epsilon}{2}$ and thus that:

$$\Delta W = F \left(-(1 - \theta)(1 - g) + \theta \frac{\lambda \epsilon}{2} \right)$$

In terms of decision-making all that matters is whether the change in social welfare is positive or negative ($\Delta W <> 0$). Since the term outside of the brackets is always

³⁵Note that this would normally be only an approximate equality but for the case of linear demand it is exact.

positive it follows that ΔW is greater than zero, and hence that marginal cost pricing delivers higher social welfare than average cost pricing, if and only if:

$$\frac{\lambda\epsilon}{2} \geq \frac{1-\theta}{\theta}(1-g) = \alpha(1-g) \quad (1)$$

In words this could be expressed as:

$$\text{‘Per-Unit’ Deadweight Loss} \geq \text{Per-Unit ‘Cost’ of the Subsidy}$$

This then is simply the original cost/benefits compared but ‘per unit’ of the fixed cost needed (F) and normalized by the appropriate distributional weights.

3.3. Empirics. This section combines estimates of the key parameters derived from the the existing literature and evidence from PSIHs with the formula derived at the end of the last section so as to provide guidance as to whether a change from an average cost to a marginal cost pricing regime would be welfare improving.

3.3.1. Distributional Weights and the Social Value of Public Funds. As previously discussed the numeraire for surplus calculations will be uncommitted government funds. It is then necessary to compute θ , the distributional weight for the specific project under consideration. Roughly the logic here is that uncommitted public funds could either be used for lowering the price of PSI or for some other government purposes. These uncommitted funds, by definition, have a weight of one.³⁶ This is almost certainly not true for the (consumer and producer) surplus generated by the project under consideration, and the appropriate distributional weight will depend on how the project’s benefits are realized across the population which in turn depends on the existing distribution of income, the degree of inequality aversion, the marginal utility of consumption and the income elasticity of demand for PSI data.

Here, it will be assumed that the *benefits from lowering the price of PSI are received in proportion to income*. Specifically, the income elasticity of consumption

³⁶That is, there is at least one project in the government portfolio where £1 of expenditure generates benefits equivalent to £1 equally distributed across the population.

of PSI is assumed to be one. This is a fair assumption given that general consumption is (approximately) proportional to income. Using this together with estimates of the distribution of income over the population and the elasticity of the marginal utility of income we obtain a range for θ of 0.718 – 0.857 with a point estimate of 0.802.³⁷

3.3.2. *Elasticity of Demand for PSI.* In this section we survey the direct and indirect evidence on the elasticity of demand public sector information and use it to form a sense of likely range of elasticities applicable in the majority of cases. The (absolute) price elasticity of demand can be interpreted as the percentage increase in demand resulting from a 1% reduction in price (or, conversely the percentage decrease in demand resulting from a 1% increase in price).³⁸ A change in pricing policy by a PSIH (or other entity) allows one to elicit the elasticity of demand by comparing prices and demands before and after the change. However, in some situations the price changes can be quite substantial. In such cases the elasticity will depend upon whether one uses the old price and output pair (before the change), or the new price and output pair (after the change). We discuss this point further in Appendix B and show that using one or other of the price-output pairs generate a set of lower/upper bounds for the elasticity.

We begin with Weiss (2004) which itself surveys a variety of existing evidence both anecdotal and systematic. He argues that the price elasticity for information is likely to be high in most cases and ‘only when use of the information is mandatory or somehow indispensable might the demand be less elastic.’

OFT (2006) also surveys existing data on the elasticities of demand for information in other countries. For example, it suggests an elasticity of 0.3 for New Zealand national mapping data based on evidence quoted by Longhorn and Blakemore (2004): ‘Rhind reviewed data charging outcomes after New Zealand had imposed a rigorous

³⁷We have omitted a detailed derivation of this figure as the calculation is a standard one and not specific to the subject matter of this paper. Those who want full details, including references for all data sources, are directed to the Appendix to Pollock et al. (2008).

³⁸Note that formally an elasticity is negative (since the price reduction is negative). However, for convenience, and to match with the definition used in the theory subsection above, the elasticity of demand has been defined so as to (normally) be positive rather than negative.

cost recovery programme on national mapping, noting a reduction in sales between 1989 and 1994 of 60%, although income was 25% greater in real terms.’ However this calculation appears to be using the lower bound – using the same calculation as in Appendix B one finds that 0.3 is the lower bound and that the upper bound is around 2.2.

Davies and Slivinski (2005) suggest that the elasticity for demand of weather forecasts is 0.3 based on evidence by Lazo and Chestnut (2002). However this paper only measures direct household demand for improving day-to-day weather forecasts through stated preference surveys. This should therefore be treated as a lower bound since it excludes demand for weather data coming from intermediaries and the private sector.

The study of Bedrijvenplatform (2000) claims ‘lowering the price of public sector geographic data by 60 per cent would lead to a 40 percent annual turnover growth’. Using this Office of Fair Trading (2006) deduce an elasticity of demand of 1.7. Interpreting turnover as revenue one finds an upper bound elasticity of 4.17 and a lower bound elasticity of 0.48 using the same calculations as in Appendix B.

Under the *Making Information Freely Available* initiative, Statistics New Zealand is in the process of making a wide range of products and data available for free.³⁹ For example, Digital Boundaries Files on CD and StreetLink files were distributed for free from July 6th 2007.⁴⁰ Digital Boundaries Files previously cost around NZ\$3,300 for the standard five-yearly census pattern, or NZ\$25,212 for the annual detailed file. StreetLink Files previously cost NZ\$6000 for first supply and then NZ\$2000 for annual updates. As of August 28th 2007 around 250 copies of Digital Boundaries CDs and 75 StreetLink files have been provided.⁴¹ This is a two-fold and ten-fold increase in Digital Boundaries Files and Street Link Files respectively in the 6 weeks

³⁹The policy press release is available at <http://www.beehive.govt.nz/Documents/Files/Statistical%20Info%20FAQ.pdf> for details.

⁴⁰See press statement available at <http://www.beehive.govt.nz/ViewDocument.aspx?DocumentID=2998>

⁴¹See statement available at <http://www.beehive.govt.nz/ViewDocument.aspx?DocumentID=30426>

after charges were withdrawn compared with what Statistics New Zealand sold in the past three and a half years.

Using these immediate changes in demand would imply very high elasticities. However the initial surge of requests could be a consequence of a backlog of demand for the data at zero-cost and so the annual uptake is likely to stabilise at a much lower level. Bearing this in mind it seems appropriate to use this recent demand to approximate the average annual uptake. Doing so and using equation (4) one finds an upper bound elasticity of around 6 and 34 for Digital Boundaries and StreetLink Files respectively. Small Area Population Estimates which previously cost around NZ\$250 were made free to download on August 28th 2007. By September 14th 2007 there had been 184 accesses by unique visitors compared to around 75 customised jobs per year previously. Again using this recent uptake to approximate the new annual output and using equation (4) one finds an upper bound elasticity of around 1.5. These estimates are still likely to be too high since the high surge in demand may include a large number of users who are unlikely to find the data of use, but request it at no cost to see if it may be suitable.

The Australian Bureau of Statistics made information free on their website towards the end of 2005. Table 6 shows the total products download statistics from 2003-2007.⁴² Figure 2 graphs usage of ABS statistics over this time period.⁴³ It is clear that there is a significant increase in the usage of data once it was made freely available. Comparing the average dissemination of 2003-2005 with 2005-07 estimates (crudely) gives an elasticity of 2.33.⁴⁴

⁴²This data is available in Table 13.3 in the ABS Annual Report at <http://www.abs.gov.au/AUSSTATS/abs@.nsf/39433889d406eeb9ca2570610019e9a5/FBF88ADA798ABCA1CA257371001411C3?opendocument>

⁴³Available at http://www.epsipus.net/content/download/7380/88070/file/3\3_ePSIplus_TM2_Pricing2_QUT_11107.pdf

⁴⁴Using the 2007 values rather than an average 2006-2007 would give an even higher elasticity of around 3.5. Thus the long-run elasticity might well be even higher – though of course one would need to then make efforts to detrend for the effect of technical advance and general growth in demand.

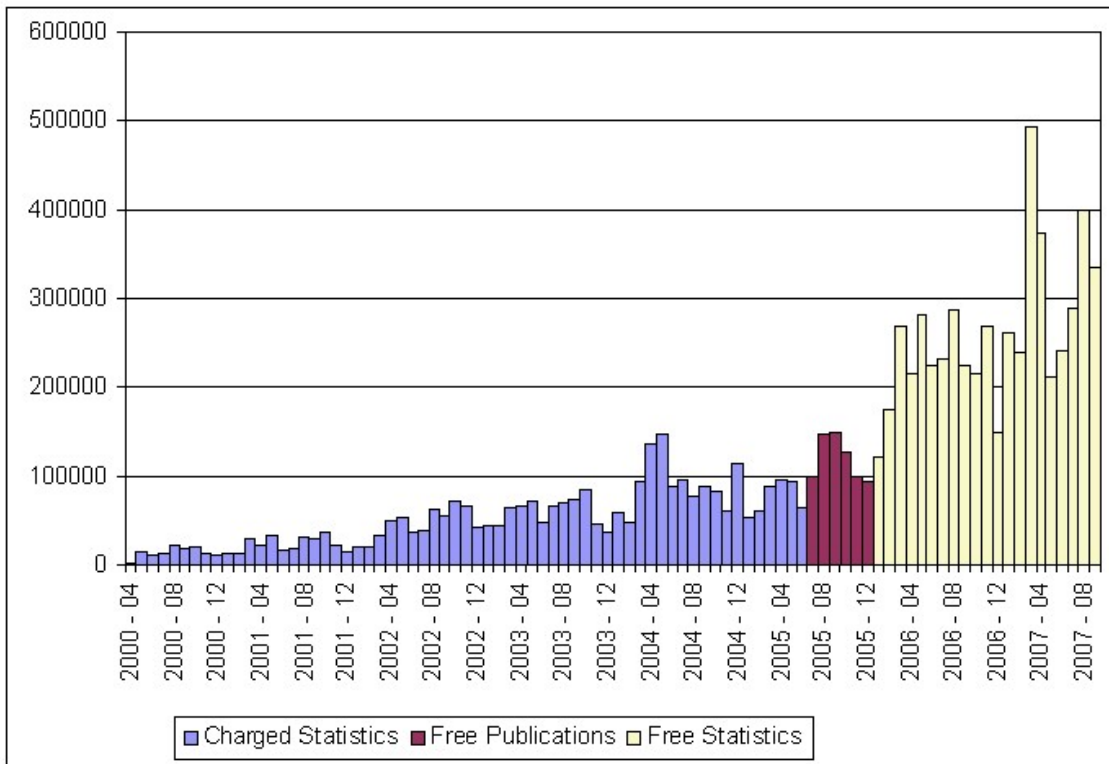


FIGURE 2. Australian Bureau of Statistics 'Dissemination of Statistics'.

	2003-04	2004-05	2005-06	2006-07
Reported	948,956	962,872	1,868,280	4,501,530

TABLE 6. Product Downloads from ABS website

The Office of Spatial Data Management in Australia conducted a wider programme to make available fundamental spatial data across a range of agencies for free or at marginal cost.⁴⁵ The policy was announced in September 2001 and implemented over a 6 month period so that by February 2002 agencies were providing data for free online, or at marginal cost in CD format. Table 7 details the delivery

⁴⁵This list of fundamental spatial datasets is listed on the Data Schedule available at http://www.osdm.gov.au/schedule/schedule_search.jsp

Year	Scheduled Dataset Units Delivered	Trend 1
2000-01	(75,310)	-
2001-02	75,310	75,310
2002-03	83,049	108,597
2003-04	52,565	156,597
2004-05	219,821	225,813
2005-06	862,530	325,622

TABLE 7. Office of Spatial Data Management Scheduled (free) Datasets Delivered. The figures in brackets are estimates. Trend 1 uses a growth rate of 44.2%.

figures for scheduled datasets (i.e. those that fell under the new policy).⁴⁶ Unfortunately no data was available for the period before the pricing policy was announced. However, if one makes the conservative estimate that data delivered in 2000-01 was no more than in 2001-02 and compares this to uptake in 2005-06 gives an elasticity of upper bound elasticity of 10.45. Of course this does not take into account any general increase in demand due to other factors. One approach would be to detrend using the ABS figures since their data did not become freely available until 2005. Using the ABS data a reasonably generous estimate for the growth rate 2001-2005 (for non-free data) would be around 44.2%. The effects of applying this growth rate is shown in Table 7 as Trend 1. Comparing the 2005-06 value in Trend 1 with the reported value suggests an elasticity of 1.65.

In order to supplement this direct evidence we also look at estimates coming from the area of telecommunications. Many analogies can be drawn between the information and telecommunications sectors making them suitable for comparison. Both are related to innovation and new technology. Both serve as inputs into other activities and both display spillover (multiplier) effects. Telecommunications is also

⁴⁶Figures from 2001-02 are quoted by OSDM as from the fundamental dataset. This is understood to be all data listed on the Data Schedule as ‘This Policy is premised on the view that all fundamental spatial data should be freely available at no more than marginal cost of transfer in order to maximise the net economic and social benefits arising from its use’ (<http://www.osdm.gov.au/fund/pricing.html>). OSDM also state that the ‘Australian government spatial datasets that are available under the terms of the Policy on Spatial Data Access and Pricing (‘the Policy’) are listed on the Schedule.’ (http://www.osdm.gov.au/schedule/schedule_search.jsp). This policy states that ‘Fundamental spatial data will be provided ... at no more than the marginal cost of transfer...’ (<http://www.osdm.gov.au/policy/accessPricing.html>)

a route through which information can be distributed and hence they are intrinsically related. The internet for example offers access to a wide range of information. Part of the demand for access to the internet will therefore reflect the demand for this information, and so the elasticities in each sector can be compared.

Hausman et al. (1997) finds a price elasticity of 1.61 and 0.51 for the introduction voice messaging and mobile phones respectively in the united states. Goolsbee (2006) finds an average price elasticity of demand for broadband of 2.75 at an average price of \$40 per month for a range of metropolitan areas in the US. Goolsbee and Klenow (2006) takes into account the opportunity cost of ones time to deduce the value of using the internet and so estimates a price elasticity of 1.6. Kridel et al. (2002) find a price elasticity of broadband of about 1.8 at \$49.95 a month. Hackl and Westlund (1996) finds a range of price elasticities of demand for international telecommunications in Sweden from 0.09 to 1.25.

To sum up, as we have just seen direct evidence on the price elasticity of information is relatively limited. Estimates often are based on large changes in prices which result in a large range of elasticities. Furthermore, the elasticity will vary depending upon the product under consideration. Thus we offer three basic ranges for the elasticity of demand for PSI which can then used as a basic for classification and discussion. The ranges are:⁴⁷

Low: 0-0.5 (midpoint: 0.25)

Medium: 0.5-1.5 (midpoint 1.0)

High: 1.5-2.5 (midpoint 2.0)

While true that the evidence is currently limited and often displays quite a range it is noteworthy that the elasticities from the direct and indirect evidence discussed above are generally quite high (i.e. greater than 1). Thus, at least with our present state of knowledge, it would appear that the medium or high range would be the most appropriate for the majority of PSI products.

⁴⁷These ranges should be interpreted as reasonably short-run elasticities. Over the long term elasticities are likely to be higher as new uses and applications for data are found.

3.3.3. *The Multiplier*. The theoretical analysis in Section 3 provided some reasonable a priori grounds for believing that the ‘multiplier’ could be significant. However, it would obviously be important to have empirical evidence for the significance and magnitude of the ‘multiplier’. Unfortunately, there is, at present, very little such evidence available. This is perhaps not surprising given the difficulties to be faced and the general lack of the detailed time-series firm-level data which would be required. However there are some suggestive individual items as well as a body of more ‘anecdotal’ evidence that can be drawn upon.⁴⁸

Weiss (2004) argues that marginal cost access to weather data in the US was a large factor in the development of the multi-billion dollar weather derivatives industry (and that its limited availability has retarded developments in the EU). An analogous argument for general weather services is made in a recent paper by Richard Pettifer, general secretary of PRIMET.⁴⁹ It argues, that particularly by comparison with the US, the EU weather marketplace is seriously underdeveloped. It goes on to argue that much of the potential, but unrealized value, lies in the ‘small unit value sector of the market place which is extremely price sensitive.’ Furthermore, and of more relevance to this subsection, realizing the potential value of those markets would involve the development of new products and services based on cheaper access to the data collected by national meteorological services.⁵⁰

⁴⁸There is, of course a significant literature on spillovers in R&D, particularly from public to private R&D. For example Jaffe (1989) and Mansfield (1995) both provide evidence of large spillover effects in this area.

⁴⁹*Towards a Stronger European Market in Applied Meteorology*. PRIMET is the association of Private Meteorological Services. Obviously, their particular interest in this area should be taken into account when considering the arguments made in the document.

⁵⁰Specifically, according to the document: “[T]his potential market [the small unit value, high potential demand] is not reached by the large government owned players because their high fixed costs and politically sponsored operating constraints prevent them from delivering the end user price and flexibility this market demands. It is not fully penetrated by the small, private sector companies largely because the exploitation of the monopoly supply position of the government owned players in respect of the raw material necessary to permit the development of suitable products at appropriate market prices. The data are subject to wholesale pricing that is too high and in some cases there is a failure to supply the data in a timely fashion (or at all), while re-use license terms can render it impossible fully to exploit the non rival nature of the data.”

Turning to geographic data, again hard data is sparse. Returning to Australia, the Spatial Information Industry Action Agenda (2001) presents evidence that reducing the price of access to geographic information had a significant impact on use and, more importantly, reuse: ‘The most important impact has been the dramatic increase in the volume of data sold. In Victoria, the number of licences or “seats” has increased from around ten before the price reductions to about 600. In Queensland, over 75 licences to distribute and value-add to the data have been issued, whereas under the previous arrangements no whole-of-state sales were made at the then commercial rate.’ Meanwhile, *Bedrijvenplatform* (2000), looking at the Netherlands suggested that a substantial portion of the benefits from cheaper geo-data would arise from the development of new products and services. In the UK, the Ordnance Survey themselves commissioned Oxera in 1999 to estimate the value of the economic infrastructure ‘built on’ OS data.⁵¹ The resulting report gave an estimate that around £79-136 Billion of Gross Value Added came from activities for which the Ordnance Survey’s geographic information was a primary input. Of course this figure does not tell one much directly about the multiplier since the fact that many businesses use (or even depend) on OS data does not itself indicate how large the spillovers are or how much innovation is occurring. Nevertheless the report is indicative of the fact that geographic information is widely used, particularly as an input into intermediate products and services, which in turn suggests the multiplier could be quite significant.⁵²

Finally, analogies can also be drawn with the spill-overs in other sectors. The Power of Information review (Mayo and Steinberg, 2007) itself provided several examples. For example, in medical studies such as Rodgers and Chen (2005) and

⁵¹See <http://www.ordnancesurvey.co.uk/oswebsite/aboutus/reports/oxera/index.html>

⁵²The argument that there are large potential gains from increased access to and reuse of PSI can be found in the PIRA report prepared for the European Commission back in 2000 (PIRA, 2000) – with a similar set of points made in OECD, Working Party on the Information Economy (Directorate for Science Technology and Industry) (2006). As with most material the contentions are based more upon analogy with the United States, and a general consideration of the market, than any ‘hard’ data – not surprising given how difficult ‘hard’ data would be to obtain.

Ziebland (2004) on breast cancer and Hellinger (2002) on HIV, it was found that access to medical information on the internet allowed users to cope better with a resulting reduction, in some cases, in treatment costs. On a different tack, Hampton (2007) finds that members of ‘wired’ neighbourhoods are more likely to know each other and Lomax (2005) finds that providing clear information with medication can improve patient adherence to medical advice by 16-33%. One could argue that similar spill-over would be present for some of the products considered below. For example, easier access to DVLA data could enable more and better HPI checks, leading to a greater return of stolen vehicles and a reduction in theft. Similarly, the Land Registry’s data on property boundaries where better access could make it easier for planners of construction projects to contact those owning neighbouring land.⁵³

Turning this diverse, and predominantly anecdotal evidence into an exact estimate for the ‘multiplier’ is clearly impossible. Furthermore, the multiplier will vary across products (just as the elasticity will). Thus, as with the elasticity, we proceed by using 3 basic ranges. Recalling that the multiplier has a lower bound of 1 corresponding to no multiplier effect on welfare, a suitable set would be:

Low (No Effect): 1

Medium: 1-3 (midpoint: 2.0)

High: 3-9 (midpoint: 6.0)

Given the great uncertainty about the exact value for the multiplier any assignment for a particular product will necessarily be substantially speculative. Thus, when performing welfare calculations, it will be important for robustness to check

⁵³One could multiply these examples of ‘potential’ applications almost indefinitely. Easier access to current and historical weather data might help those researching climate change. Better access to geographic information would enable greater citizen understanding and participation in the planning at the local, regional and national level. Increased freedom of reuse would greatly multiply the potential for specific groups, whether those with disabilities such as the blind or with particular interests such as walkers, to add value to basic geographic data whether via annotation or integration with other sources of data.

all results using a multiplier of one (i.e. no effect). This way, while the multiplier is incorporated into the analysis one can also be sure that it is not ‘driving the results’.

3.4. Summary and Conclusion. As should be clear from the preceding discussion, while our estimate for the distributional weight θ is fairly good, our estimates of the other major parameters are highly uncertain. However, recall from equation (1), that a MC pricing regime is equal to or superior to AC if and only if:

$$\frac{\lambda\epsilon}{2} \geq \frac{1-\theta}{\theta}(1-g)$$

Substituting $\theta = 0.802$ and multiplying both sides by 2 gives:

$$\lambda\epsilon \geq \frac{1-g}{2}$$

Assuming the proportion of government usage is ≈ 0 (the most pessimistic scenario for the marginal cost regime) then gives that MC pricing is superior to AC pricing if:

$$\lambda\epsilon \geq \frac{1}{2}$$

3.4.1. Charging for Use. Thus, deciding on the charging regime for a given piece of PSI reduces to deciding whether the product of the multiplier and the elasticity is greater than a half. Since the multiplier, λ , is always greater than or equal to 1, an elasticity above a half is sufficient to imply that marginal cost pricing is preferable to average cost pricing.⁵⁴ The full sets of outcomes as a function of the 3 categories (low, medium, high) for the multiplier and the elasticity are shown in Table 8.

The evidence presented previously suggest that for most examples of (digital, upstream) public sector information the medium or high range for the elasticity would be the most appropriate. Thus, for these kinds of public sector information **marginal cost pricing for users is the preferable option.**

⁵⁴Though, as discussed previously, we are likely to be dealing with large changes in price and demand. In this case we might want to be cautious, particularly given the underlying linearity assumption, and require a somewhat higher elasticity and/or multiplier before being certain that marginal cost were preferable.

		λ		
		1	1-3	3-9
ϵ	0.0-0.5	AC	AC/MC (AC)	AC/MC (MC)
	0.5-1.5	MC	MC	MC
	1.5-2.5	MC	MC	MC

TABLE 8. Preferred charging regimes for different parameter ranges. AC (MC) indicates that average cost (marginal cost) pricing is preferable throughout the range. AC/MC indicates that average cost and marginal cost pricing is preferable, but in different parts of the range (the figure in brackets indicates what is preferable at the mid-points of both ranges)

3.4.2. *Charging for Updates.* So far in our discussion of funding policies we have large ignored the possibility of ‘updater’ or ‘write’ funding, that is charging those who make updates to the information set. Though in some cases there are no ‘updaters’ to charge in other cases, perhaps the majority, there clearly are – in fact some PSIH’s primary function is ‘registration’.⁵⁵

Importantly all of our theoretical analysis developed for normal information ‘users’ can also be applied to the case of ‘updates’. That is, just as for ‘users’, we can compare charging ‘updaters’ a price equal to the marginal cost of their activities or a price equal to average cost (that is sufficient to cover fixed costs as well). Since the analysis is essentially identical all we need to do is replace the elasticity of demand with the elasticity of registration/updates and the multiplier for use with a ‘multiplier’ for updates.⁵⁶

⁵⁵For example, PSIHs dealing with property, vehicle and company ownership. Non-registration examples are perhaps more interesting here because less obvious. One important example relates to geospatial datasets. Today, many of the changes which occur to such datasets are likely to be anthropogenic, arising from activities such as construction, road-building etc. As such, it would be possible to levy charges upon those carrying out the activities leading to changes in a geospatial dataset (and such changes are often already being logged via some form of planning process).

⁵⁶In fact we need to a little bit more: for ‘updates’ we cannot take marginal cost as zero and hence the formula would become: $\lambda\epsilon \geq \frac{\Delta p}{2p}$ where Δp is the reduction in price when going from average cost pricing (p) to marginal cost pricing. We might also want to perform a more complete two-sided analysis. This would involve taking account of the impact of each side on the other – i.e. the impact of more ‘updates’ on the demand of ‘users’ and of an increase in ‘users’ on the demand for ‘updates’. However, for most public sector information it is reasonable to assume these effects are fairly negligible (‘updaters’ especially care little about ‘users’ – though an important exception is provided by the case of harbours which often need to be properly charted to be usable by marine vessels).

This is instructive because the parameters for ‘updates’ are likely to differ substantially from those for ‘use’. In particular, for ‘updates’ there are no systematic distortions of the willingness-to-pay. As such multiplier effects are likely to be negligible ($\lambda = 1$). Furthermore, the elasticity of ‘demand’ for updates is likely to be (very) low: most such updates occur as a very small, but legally required part, of some larger activity – such as buying a house, creating and running a limited company, etc. Thus, contrary to the situation on the ‘user’ side, charges for ‘updates’ above marginal cost may well be an attractive option, and preferable to funding a PSIH’s fixed costs out of general government revenues.

4. CONCLUSION

This paper has provided an overview of the economics of public sector information with particular attention to the regulatory structure and funding model under which public sector information should be collected, maintained and provided. The focus was on public sector information that was *digital, non-personal and upstream*.

Funding for public sector information can come from three basic sources: government, ‘updaters’ (those who update or register information) and ‘users’ (those who want to access and use it). Policy-makers control the funding model by setting charges to external groups (‘updaters’ or ‘users’) and committing to make up any shortfall (or receive any surplus) that results. Much of the debate focuses on whether ‘users’ should pay charges sufficient to cover most costs (average cost pricing) or whether they should be given marginal cost access – which equates to free when the information is digital. However, this should not lead us to neglect the third source of funding via charges for ‘updates’.

Policy-makers must also concern themselves with the regulatory structure in which public sector information holders operate. The need to provide government funding can raise major commitment questions while the fact that many public sector information holders are the sole source of the information they supply raise serious competition and efficiency issues.

Having an adequate structure in place is essential for a public sector information holder's performance – be that in terms of efficiency, information quality etc. Furthermore, such a structure is important under all funding options – and this structure should (and can) be chosen independently of that funding option. Thus, getting this right must be one of the first items on the agenda of any policy-maker concerned with the operation of public sector information holders.

Regulation should be transparent, independent and empowered. For every public sector information holder there should be a single, clear, source of regulatory authority and responsibility, and this ‘regulator’ should be largely independent of government. Policy-makers around the world have had substantial experience in recent years with designing these kinds of regulatory systems. As such, this is not a task one that should be especially difficult to complete.

Turning to the question of funding, there is a general proposition that public sector goods and services should be offered at efficient prices, unless there are compelling reasons to depart from efficiency. In the absence of beneficial (or harmful) spillovers, the efficient price is marginal cost (with supply adapted such that the short and long-run marginal costs are equal). One reason for departing from efficient pricing is that the marginal cost is below the average cost, and that the benefits of a hard budget constraint outweigh the distortionary costs of raising the revenue to make up the short-fall, not from general taxation, but from raising the price of the products supplied.

When it comes to charging ‘users’ of public sector information the case for pricing at marginal cost or below is very strong for a number of complementary reasons (note that, for most digital data, marginal cost will be approximately zero). First, the distortionary costs of average rather than marginal cost pricing are likely to be high because: a) the mark-up to cover fixed costs is high, as marginal costs are such a low fraction of average costs b) the demand

for digital data as with other information services is likely to be high and growing c) there are likely to be large beneficial spill-overs in inducing users to innovate new services based on the data, as is evidently the case for other ICT services. Second, the case for hard budget constraints to ensure efficient provision and induce innovative product development is weak for public enterprises not subject to regulation and providing monopoly services without fear of competition. It would be far better to address issues of incentives, regulation and commitment explicitly rather than indirectly through budget constraints. Finally, for several services, the government is already providing effectively a large contribution to fixed costs, without allowing the public to enjoy the benefits of efficient pricing.

By contrast, it may well be good policy to charge ‘updaters’ of public sector information prices above marginal costs, using the funds thereby obtained to cover (some portion) of the fixed costs of maintenance and collection. This is in accordance with good Ramsey pricing principles that if distortionary mark-ups are necessary to cover or contribute to fixed costs, they should be higher for inelastically demanded goods and lower for elastically demanded ones (in simple cases, the mark-up divided by the price should be inversely proportional to the elasticity of demand).

To conclude: **most upstream, digital public sector information is best funded out of a combination of ‘updater’ fees and direct government contributions with access provided for free. Appropriately managed and regulated this model offers major societal benefits from increased provision and access to information-based services while imposing a very limited funding burden upon government.**

APPENDIX A. THE DEMAND CURVE AND SOCIAL WELFARE

A.1. **Intermediate Firms.** Consider, as an example, the case where a PSIH sells to a downstream firm which is a monopolist in its own market. In that case, with royalty-based pricing, one would have a classic case of ‘Cournot complements’ and

attendant double marginalization, and the demand curve seen by the PSIH would under-represent actual demand and welfare changes.⁵⁷

A similar, but different, effect arises if downstream firms have fixed costs as a result of the Dupuit triangle.⁵⁸ Imagine there are a large number of downstream firms each demanding one unit of the information but with different fixed costs. The PSIH's demand curve then arises from aggregating across all these downstream firms. Pick a point on the PSIHs demand curve, p, q say, and consider an increase of δp in the price charged resulting in some reduction δq in purchases.⁵⁹

This reduction in demand corresponds to some downstream firms ceasing to purchase (and hence ceasing production). Consider one of these firms and let initial revenue be R and C their total costs (excluding the payment for data). Then one must have $R - C \approx p$ (since $R - C < p + \delta p$ and $R - C \geq p$). What about the surplus generated by this firm? Its producer surplus is zero ($R - C - p = 0$) but consumer surplus, denoted CS , is almost certainly not zero. Thus, from the point of view of society current total surplus produced by this firm is $p + CS$. However using the demand curve of the PSIH all that would be recorded is the p coming from the payment for data.⁶⁰

A.2. Dynamics and Innovation. Lower prices for information today, by increasing access and usage, are likely to stimulate the rate of innovation both downstream and in related (esp. complementary) markets. For example, cheaper geodata may lead to more rapid improvement in the quality of the software and hardware components of Geographical Information Systems (GIS). Or, as Weiss (2004) argued,

⁵⁷Note that this effect still occurs if the downstream market is an oligopoly rather than a monopoly though the degree of double-marginalization will decrease as the level of competition increases.

⁵⁸This effect occurs whether the tariff used by a PSIH is a royalty or a fixed fee – unlike the case of ‘Cournot complements’.

⁵⁹Since only one unit of the product is demanded here this is necessarily a fixed fee. However this argument can easily be extended to the more general case where demand is variable and the PSIH sets a nonlinear tariff.

⁶⁰This, of course is in the extreme case where the firms who no longer purchase simply cease operation. However the basic point still holds in the more realistic case where a rise in the price of the PSIH's product causes them to substitute it with another (necessarily inferior) input.

marginal cost access to weather data in the US was a large factor in the development of the multi-billion dollar weather derivatives industry.

It is quite possible for such effects on welfare to be large, much larger in fact than those arising from purely static considerations related to the underlying product's demand curve.⁶¹ Furthermore this may be true even if the current costs of access are relatively low – at least relative to the potential benefits. This is somewhat surprising since normally one would imagine that the cost of a particular piece of information should place a (rough) upper limit on the value of the innovations which it enables.⁶²

However, there are a variety reasons why this basic logic fails. The simplest example is to consider a chain of cumulative innovations in which an innovator at each stage can only extract some fraction r of the total surplus generated by the subsequent innovator. In this case for a chain of length N the initial innovator only receives r^N of the actual surplus generated (and so conversely an innovator with a willingness to pay of only X for a piece of data may be generating a surplus of X/r^N).

Another possibility is that the innovation effort is distributed across many different firms or individuals ('componentized' innovation – as an explicit example one could think of an open-source project working to produce GIS software). In this case if each agent needs access to the underlying data supplied by the PSIH in order to contribute to the project the total costs may become so high as to be prohibitive.⁶³

⁶¹Of course, in doing such a calculation, one would need to be cautious about how one allocated these 'spillover' benefits. Just because the data provided by a PSIH is used in (or is even central to) the activities of a particular firm does not mean one can allocate all the surplus generated to the availability of that data.

⁶²To give a concrete illustration, suppose a particular set of geodata costs £1000 and there is a potential innovator who has an idea for a new product based on that geodata worth £ X . If $X > 1000$ then the innovator should be willing to pay for access to the geodata. This suggests that only innovations worth less than £1000 are lost when the price is at this level.

⁶³Suppose the innovation is worth V and the cost of data is X and there are N participants. If these N participants were all in a single firm which could obtain a single development license the cost is X , willingness to pay is V and the project is undertaken if $V \geq X$. However if the participants are distributed and must all buy their own license then the willingness to pay of any individual would be only V/N and the project is only undertaken if $NX \leq V$.

Other more complex examples can be found in the cumulative innovation literature – see e.g. Bessen and Hunt (2007); Pollock (2008).

APPENDIX B. DERIVING THE ELASTICITY OF DEMAND

The formal definition of price elasticity of demand is given in the expression below where p is the price of data, q is output and where δ represents an infinitesimal increase in the variables:⁶⁴

$$\epsilon = -\frac{(\delta q/q)}{(\delta p/p)}$$

Intuitively one can think of the price elasticity of demand as the percentage increase in demand for PSIH data for a one percentage point decrease in price. Similarly, a change from average cost to marginal cost pricing (or vice-versa) allows one to elicit the elasticity of demand. However as the price changes can be quite substantial, the elasticity will depend upon whether one uses the old price and output pair, p_0, q_0 , or the new price and output pair, p_1, q_1 . This is best illustrated with an example, say where a price rise results in revenues increasing by 40 percent and output decreasing by 40 percent. Let R_0 and R_1 be old and new revenues respectively. The relation between old and new revenues and old and new prices can be expressed as:

$$R_1 = \frac{7}{5}R_0 \tag{2}$$

$$q_1 = \frac{3}{5}q_0 \tag{3}$$

Now assume that the demand curve is linear. If one defines the elasticity using the old price output pair then:

$$\epsilon_0 = -\frac{(q_1 - q_0/q_0)}{(p_1 - p_0/p_0)} \tag{4}$$

⁶⁴Note that usually there would be no negative sign at the front here. However, for convenience, and to match with the definition used in the theory subsection above, the elasticity of demand has been defined so as to (normally) be positive rather than negative.

$$R_1 = p_1 q_1 = \frac{7}{5} p_1 q_1 = R_0 \quad (5)$$

Substituting (3) into the (5) gives:

$$p_1 = \frac{7}{3} p_0 \quad (6)$$

Similarly substituting (6) into (4) one finds that $\epsilon_0 = - - 3/10 = 0.3$. However if one defines the elasticity using the new price output pair then:

$$\epsilon_1 = - \frac{(q_1 - q_0/q_1)}{(p_1 - p_0/p_1)} \quad (7)$$

Substituting (6) into (7) one finds that $\epsilon_1 = - - 7/6 = 1.17$.

For a linear demand curve the high price, low output pair generates a significantly higher elasticity. Note also that the demand curve may not be linear. Demand may be more inelastic at higher prices, where there are a few large businesses who simply have to have the data and so are willing to pay a very high price. However the demand may also be particularly elastic lower down the demand curve, where a substantial amount of experimentation with the data may take place. Both these effects would reduce the effect on elasticity estimates with different price, output pairs. Nonetheless, wherever there are substantial price changes, there will also be a significant range in elasticity estimates using the same underlying data. In such cases it may be more appropriate to use the mid-point as opposed to either the upper or lower bounds.

REFERENCES

- Bedrijvenplatform, R. (2000). Economische effecten van laagdrempelige beschikbaarstelling van overheidsinformatie. Private sector members of the Dutch Geographic Data Committee.
- Bessen, J. and Hunt, R. M. (2007). An Empirical Look at Software Patents. *Journal of Economics & Management Strategy*, 16(1):157–189.

- Davies, J. B. and Slivinski, A. (2005). The Public Role in Provision of Scientific Information: An Economic Approach. University of Western Ontario, RBC Financial Group Economic Policy Research Institute Working Papers 20051, University of Western Ontario, RBC Financial Group Economic Policy Research Institute.
- Farrell, J. (2003). Integration and Independent Innovation on a Network. *AER Papers and Proceedings*, pages 420–424.
- Goolsbee, A. (2006). The Value of Broadband and the Deadweight Loss of Taxing New Technology. *Contributions to Economic Analysis & Policy*, 5(1):1505–1505.
- Goolsbee, A. and Klenow, P. J. (2006). Valuing Consumer Products by the Time Spent Using Them: An Application to the Internet. *American Economic Review*, 96(2):108–113.
- Hackl, P. and Westlund, A. H. (1996). Demand for international telecommunication time-varying price elasticity. *Journal of Econometrics*, 70(1):243–260.
- Hampton, K. N. (2007). Neighborhoods in the Network Society: the e-Neighbors Study. In *Information, Communication and Society*.
- Hausman, J. A. (1981). Exact Consumer's Surplus and Deadweight Loss. *The American Economic Review*, 71:662–676.
- Hausman, J. A., Pakes, A., and Rosston, G. L. (1997). Valuing the Effect of Regulation on New Services in Telecommunications. *Brookings Papers on Economic Activity. Microeconomics*, 1997:1–54.
- Hellinger, F. J. (2002). Focus on Research: HIV Disease. Technical report, Agency for Healthcare Research and Quality, United States Department of Health and Human Services.
- Jaffe, A. (1989). Real Effects of Academic Research. *American Economic Review*, 79:957–970.
- Kridel, D., Rappoport, P., and Taylor, L. (2002). The Demand for High-Speed Access to the Internet: The Case of Cable Modems. In Loomis, D. and L.D. Taylor, K., editors, *Forecasting the Internet: Understanding the Explosive Growth of Data*

- Communications*. Academic Publishers, Dordrecht.
- Lazo, J. K. and Chestnut, L. G. (2002). Economic Value of Current and Improved Weather Forecasts in the US Household Sector. In *Report prepared for the National Oceanic and Atmospheric Administration by Stratus Consulting Inc. Boulder, Colorado*.
- Lomax, D. G. (2005). Presentation to Patient Compliance, Adherence and Persistence Conference.
- Longhorn, R. and Blakemore, M. (2004). Re-visiting the Valuing and Pricing of Digital Geographic Information. *Journal of Digital Information*, 4(2).
- Mansfield, E. (1995). Academic Research Underlying Industrial Innovations: Sources, Characteristics and Financing. *Review of Economics and Statistics*, 77:55–65.
- Mayo, E. and Steinberg, T. (2007). The Power of Information.
- OECD, Working Party on the Information Economy (Directorate for Science Technology and Industry) (2006). Digital Broadband Content: Public Sector Information and Content. Technical report.
- Office of Fair Trading (2006). The Commercial Use of Public Information.
- OFT (2006). Commercial Use of Public Information; Annexe G: Economic value and detriment analysis.
- PIRA (2000). Commercial Exploitation of Europe’s Public Sector Information. Final Report for the European Commission, Directorate General for the Information Society. Technical report.
- Pollock, R. (2008). Cumulative innovation, experimentation and the hold-up problem. Cambridge Working Papers in Economics 0817, Faculty of Economics, University of Cambridge.
- Pollock, R., Newbery, D., and Bently, L. (2008). Models of Public Sector Information Provision via Trading Funds. Published by BERR (commissioned by HM Treasury and BERR).

- Rodgers, S. and Chen, Q. (2005). Internet Community Group Participation: Psychosocial Benefits for Women with Breast Cancer. *Journal of Computer-Mediated Communication*, 10(4).
- Spatial Information Industry Action Agenda (2001). Positioning for Growth. Technical report.
- Weiss, P. (2004). Borders in Cyberspace: Conflicting Government Information Policies and their Economic Impact. In Aicholzer, G. and H., B., editors, *Public Sector Information in the Digital Age: Between Markets, Public Management and Citizens' Rights*.
- Willig, R. D. (1976). Consumer's Surplus Without Apology. *The American Economic Review*, 66:589–597.
- Ziebland, S. (2004). The Importance of Being Expert: the Quest for Cancer Information on the Internet. *Social Science and Medicine*, 9(59):1783–1793.