

APPSI RESPONSE TO THE PUBLIC DATA CORPORATION (PDC) CONSULTATION

Introduction

This statement forms a response to the publication of a public consultation: ***Data Policy for a Public Data Corporation***. A separate response has been provided to the partner consultation: *Making Open Data Real*.

We begin with a high level statement designed to identify and address some big PDC-related issues for the forthcoming White Paper which will set the framework for a PDC. This is followed by specific answers to the direct questions posed in the consultation paper where we are able to answer them.

APPSI is an independent body¹ established by the UK government to provide advice to Ministers and to The National Archives and the Controller of Her Majesty's Stationery Office. It is also part of the statutory Appeal Review regarding the re-use of Public Sector Information Regulations (Statutory Instrument 2005 No. 1515). Its members have a rare level of insight regarding PSI, consisting of highly experienced specialists with backgrounds in the civil service, local government, health, trading funds, academia, private sector (users, developers and consultants) and the third sector. All countries in the UK are represented. Skills are present in policy, technical, legal and economic areas². All members of APPSI took part in discussions on this topic, declaring any possible conflicts as appropriate. The result does not necessarily reflect the views of each and every member on every topic but considerable consensus was achieved.

APPSI members will be pleased to support the ongoing development and implementation of the policies.

SECTION A: SUMMARY OF APPSI POSITION ON THE PDC

APPSI welcomes government's continued endeavours to address the challenges and seize the opportunities associated with open data; the UK has made much progress in the recent past. We have though had considerable difficulty dealing with the PDC consultation. In part this was because we saw the concept of the PDC articulated in this consultation and in press statements being unclear: it has been characterised by some as a solution looking for a problem. It certainly focuses on information trading much more than Open Data.

After much discussion, we believe that the key issues are ones of data and public policy more generally than is addressed in the consultation document. These are manifested as the deleterious effects of the Trading Fund model, the inchoate state in which multiple related UK policies and statutes have evolved in isolation, and the crucial need for sustainable financing of core national databases. Thus we agree strongly with – and would go further than – the government statement in the Open Data Consultation (page 14) that: "government's approach to the release of public data can be described as haphazard and in need of reform".

¹ <http://www.nationalarchives.gov.uk/appsi/default.htm>

² One of APPSI's members is employed by one of the Trading Funds being considered for membership of the PDC. In these circumstances he is conflicted and cannot comment on any of the matters contained herein.

The Public Data Corporation consultation has however served as a trigger for APPSI re-thinking a more radical and strategic approach³. Without such an approach, we expect that the existing complexity, lack of clarity and tensions will continue. We set out below what changes we think are necessary, why they are needed and how best they can be affected. Within this we set out some ideas for a different model of a PDC.

The stated objectives of the PDC⁴

These are to:

- provide a more consistent approach towards access to and accessibility of public sector information, balancing the desire for more data free for re-use whilst ensuring affordability and value for taxpayers;
- create a centre of excellence driving further efficiencies in the public sector; and
- create a vehicle that can attract private investment.

We admire these objectives but believe that they are impossible to achieve under present circumstances and plans because of a number of deep-seated problems.

Shortcomings of the current PDC proposal

We see these as:

- The consultation proposal conflates mutually incompatible functions together: operations, sales and marketing, standards setting, procurement, policy-making, and a vehicle for private investment;
- The latter part of the first objective (i.e. "...balancing the desire for more data free for re-use whilst ensuring affordability and value for taxpayers") is an aspiration rather than an objective, there are no proposed quantitative measures of PDC success and no business case. Such measures might include providing widely used and successful 'data discovery' tools;
- It appears to create complexity (e.g. with two levels of Board involved and brings together (possibly) three or more organisations which presently address very different markets, have different funding streams (mostly from government bodies for the Met Office and Ordnance Survey) and have different users. The governance structures are not clear – which impacts on transparency and on public acceptability;
- As we understand it, the Public Sector Mapping Agreement has set the framework for much of Ordnance Survey's activities for some years to come and the Met Office is tied into international agreements for data sharing and pricing;
- Even if the PDC consisted solely of the three organisations transferred to BIS, the combined turnover would be around £700m. This huge entity could well create real competition issues if it effectively became both data gatekeeper and principal competitor to (probably much smaller) private sector competitors;
- If the PDC was a vehicle that sought to attract private investment, it would either have to extract monopoly rent from particular datasets or go beyond the PDC's public task to compete with existing value-added resellers. This seems unwise in a government-controlled organisation (as we understand it to be).

³ We readily understand the constraints of public finances in bringing about significant change. It should be clear that APPSI members are not opposed to charging for some types of information (see our response to the Open Data consultation) providing this is part of a coherent policy and the charging is fair and equitable.

⁴ Source: page 11 of consultation document

THE BIG PROBLEMS

Aside from our specific concerns with the PDC as currently proposed, we identify three main problems with the current situation. These are briefly described below with a proposed way forward; elements of our argument are included in relevant responses to consultation questions.

The Trading Fund model

This consultation is very largely about three organisations which are data rich and are Trading Funds. Many other data-rich organisations exist in government which are not Trading Funds but some have a history of charging for ‘their’ data (e.g. the Environment Agency). This does not appear to have been covered in any significant way in the consultation.

Trading Fund status has been bestowed on these three (and some other) organisations at different times and for different reasons. Each was adopted out of expediency. If there is a coherent underpinning concept it is that these government bodies must be able to support their data collection, collation and creation activities through trading in products and services and use the proceeds of that trade to generate a return on capital. Trading Fund status has focused the Funds on delivering what some customers want to a greater extent than was the case previously. But other consequences of this situation are that the Trading Funds:

- Have every incentive to protect their IPR and maximise revenues from them. Given the fact that data from the Land Registry, Met Office and Ordnance Survey are essential underpinnings of the use of much other data, this has a direct effect on the entire success of the Open Data programme;⁵
- Have every incentive to maintain current revenue and continue with an existing business model rather than seek to grow volume and reduce unit costs to users (which would pose larger risks to their finances);
- Engage in direct competition with private sector bodies (which has led to accusations of unfair practices and the recognition that tighter regulation is essential. This is exacerbated through resellers having to expose their business plans to the Trading Fund before permission to the data is granted.
- Take a narrow view of the benefits and public good arising from their information assets – in contrast to the much wider view taken by government in the Open Data programme and consultation.

These consequences of the Trading Fund model are not intended as gratuitous criticisms – they are inevitable results of the model – but are inimical to the aims of Open Data and generate unhelpful behaviours in the public and private sectors.

The confusion of legislation and guidance

These and other organisations trading in data and information are faced with a cat’s cradle of legislation and formal guidance, including but not restricted to:

- Crown copyright and Crown database rights
- The Information Fair Trader Scheme

⁵ The situation has recently been eased somewhat so far as Ordnance Survey has been concerned in that significant amounts of OS data have been made freely available for re-use by all and that other data has been made available free at the point of use within the public sector. The private sector is required to pay for OS Master Map and some other products. As a result, Ordnance Survey now operates under a hybrid partly government-funded, partly commercial model

- The Freedom of Information Act(s)
- Rights of Privacy arising under the Human Rights Act and common law
- The Data Protection Act
- Re-use of Public Sector Information Regulations
- Environmental Information Regulations
- The INSPIRE Regulations
- Competition law
- 'Managing Public Money' guidance from HM Treasury
- Local government regulation (e.g. the powers of charging for discretionary services under the Local Government Act 2003)

This legislation and guidance has been promulgated at different times, for different purposes and under different imperatives. There has been no holistic thinking about what will achieve contemporary policy objectives. Some of the texts are also contradictory e.g. a different charging basis for re-use in the Environmental Information Regulations versus other legislation; conflicts between the rules on charging in the Re-Use of Public Sector Information Regulations and in regulations affecting local authorities; and can be difficult even for lawyers to interpret (e.g. Regulation 15 of the Re-Use of Public Sector Information Regulations). Some of it is interdependent (e.g. the need to make changes to the local government Land Charges regulation in response to complaints to the Information Commissioner under the Environmental Information Regulations⁶).

Some of this legislation originates at European level. We would echo here Professor Hargreaves' recent call in his review⁷ (accepted by the Government in its response⁸) for the UK Government to exercise influence in shaping policy and legislation at international level, based on objective economic evidence. The same should of course inform policy at national level.

Such is the complexity of the legislative framework and consequent friction impeding success of the Open Data programme that we believe it is now time for fundamental reform. As a first step we need a detailed synthesis of the conflicts in the relevant legislation.

Financing

We recognise the need for sound financing of information infrastructure. We also recognise the folly of simply mandating that all PSI must be made available for re-use without making any consideration of funding; local government has been particularly vociferous about this point but we see the same views in the NHS and indeed in Whitehall government. In some cases governments have funded the data creation in its entirety. The big issue if the taxpayer is not providing the funding is who should pay, when to charge and on what basis – and what will be the consequence for sustainability of data supply and quality.

Multiple financing models exist and have been used in many sectors. Just as in physical infrastructure (e.g. toll bridges), funding by users has attractions compared to taxpayer funding. The consultation hypothesises the merits of Freemium mechanisms. We see tactical devices like price-discovery systems as being useful approaches, where pricing for information can sometimes be optimal by discriminating across users. Whether this is acceptable in a government environment is not clear. Based on experience in the transport sector, we also have reservations about the extent to which such approaches can be effective across the PSI landscape and are sustainable over time. The essence of many parts of our information infrastructure is that they must be predictably and regularly up-dated and maintained. We note

⁶ See : <http://www.legislation.gov.uk/uksi/2010/1812/memorandum/contents>

⁷ 'Digital Opportunity – A Review of Intellectual Property and growth', May 2011: see <http://www.ipo.gov.uk/ipreview-finalreport.pdf>.

⁸ See: <http://www.ipo.gov.uk/ipresponse-full.pdf>.

that the maintenance even of popular data sets⁹ tends to degrade when initial funding diminishes. For example, Neighbourhood Statistics produced by ONS was called into being with significant initial funding to meet information needs identified in a Cabinet Office study but has been little enhanced in subsequent years. The charging of user fees for services based on access to a database had its attractions but the experience of HM Land Registry – where income dropped by a third in the three years from 2008/09 – is salutary. Another often advocated alternative is marginal cost pricing for government data.

In recent Cabinet Office workshops, few from the business community present opposed some charging, saying that speed and ease of access to data, good access to civil servants familiar with its data details, licensing and other factors were more important than questioning the principle of charging itself¹⁰. Government might significantly reduce the difficulties encountered in re-using PSI by aiding public bodies to identify assets, put upon them a timely obligation to respond to re-use requests, and create a more empowered/resourced mediation process than exists at present.

What we do know is that where data are free at the point of use, the scale of use often expands very rapidly – but this may only be true of some key data sets. For example, before 2001 the Population Census statistics were charged for at a non-marginal price and were arguably underutilised across government and the private sector. This changed to ‘free at the point of use’ with the Census Access initiative (using the Invest to Save Budget) and the effect was dramatic. The statistics became very widely used, much better investment and service provision resulted and a whole new sector of commercial exploitation opened up, with employment and taxable revenue gains. User needs became better understood and the current Census was of better quality for that. In addition, the view of many in APPSI is that the taxpayer-funded OS Open Data and the Public Sector Mapping Agreement had a very positive impact on the exploitation of Ordnance Survey data.

What follows from all of this is that, to get the benefits of wider use, the sustainability of the needed quality of ‘core databases’ is of vital importance and the funding structure needs to ensure that this exists: this suggests that some form of continuing government funding in – and ownership of - those databases deemed critical for the National Information Infrastructure (Annex 1) may be necessary but that individual products could be funded on a different model.

⁹ APPSI members are clear that the National Information Infrastructure is largely concerned with core databases rather than products spun off from them.

¹⁰ This differs from the public reaction during the Guardian *Free our Data* campaign. We should also note that capacity and willingness to pay differs hugely between different types of firms (e.g. SMEs as opposed to major utility companies or insurance firms). We are also aware that the Professional and Business Services Group, Chaired by Sir Michael Snyder, and comprising leading UK business service firms such as PwC, KPMG, Accenture, and Arups is concerned that charging for government data is holding back business exploitation; it is recommending a radical new approach by establishing the UK as a creative commons and introducing more powerful/transparent methods of stakeholder engagement into the governance of data access.

OUR CONCLUSIONS ON HOW TO GO FORWARD

The expediency of the Trading Fund model, the lack of any coherent metric for assessing total costs and benefits from it¹¹, the large number of policy players involved¹², the difficulty associated with partitioning activities¹³ and costs and the variations in terminology used combine to create a situation which needs completely re-thinking.

We argue that government needs to:

- Provide clearer leadership, notably in providing a new, consolidated statutory framework embracing the whole public sector. This will require new legislation¹⁴ and a strong regulator;
- Be much more strategic in deciding:
 - *Which part of government has lead responsibility for Public Data / PSI, its production, access to it and its re-use?*
 - *What do governments and the wider public sector need by way of data and information to fulfil their responsibilities and obligations, including meeting the 'public good', i.e. ensuring wider societal benefits through the delivery of public services and PSI? This necessitates the definition of a contemporary National Information Infrastructure (see Annex 1). In essence, some data are more important than others. The priority should be the most important ones required to run the country, permit the public to monitor government performance and public probity and stimulate innovation and growth¹⁵. This requires data curatorship, good governance and strategic development of these data by governments;*
 - *What are the respective roles of government(s) and of the private sector in providing the necessary data and information, then maintaining it at a suitable level of quality and currency? So far as the public sector is concerned this amounts to having a strategic view of the public task, rather than versions crafted for each organisation by its own hand¹⁶;*
 - *What charging policy should be applied to different classes of data and why? APPSI members are clear that access to 'Core Reference Datasets' – those where there is a moral right for the public to have access (e.g. details of all the UK laws) or where there are large safety, public service, efficiency and cost benefits if everyone uses the same definitive and regularly up-dated sources of data – should be available at no*

¹¹ This is not for want of effort: there have been many studies of the economic benefits of trading government data and the size of the putative PSI marketplace since even before Shapiro and Varian's classic book. One recent EU-sponsored study is summarised in *Review of recent studies on PSI Re-Use and related market developments* by Vickery G (2011) see <http://epsiplatform.eu/content/review-recent-psi-re-use-studies-published>. We also are aware that Cabinet Office is participating in a current study on the benefits of having made some OS data freely available. We find many of these interesting but it is inescapable that all rely on complex and often un-testable assumptions.

¹² Including the Cabinet Office, the Ministry of Justice, The National Archives, the Shareholder Executive and BIS, and of course the devolved administrations. Other organisations such as the Information Commissioner's Office, ONS and the NHS have influence on what emerges.

¹³ Such as what is 'value added' c.f. 'raw data' except in a few obvious cases

¹⁴ There is a precedent for such a 'Public Data Act'. The Statistics and Registration Service Act 2007 was the first major legislation on official statistics for 60 years. It transformed the statistical landscape by setting the objective of having high quality and trustworthy statistics, bestowed significant independence on the UK Statistics Authority (which answers to Parliament) and defined the roles of the Office for National Statistics, the National Statistician and the Head of Assessment (the 'quality inspector') as key players in the new arrangements. Legislation requirements will need to be kept under review, depending upon the outcome of the review of the EU PSI Directive.

¹⁵ We note that www.data.gov.uk, which has access to 7000 data sets, has been an outstanding technical success but the economic and social value is at present wholly unknown.

¹⁶ We can see at least one case where an added value role for the public sector may be inescapable. Where individual data are collected under statute but-purpose-specific outputs are desirable (e.g. income data for geographical areas achieved through aggregation of individual tax records).

more than marginal cost. In an Internet world this translates to virtually free at the point of use. In our submission to the Open Data consultation, we would say that we expect that all 'exhaust data' – that produced as a by-product of government activities for internal purposes – would be made available at no cost to users, subject to usual safeguards, and provided (as will normally be the case) that there are only very modest costs of so doing.

- *What are the appropriate roles of the public and the private sectors in disseminating these data and information?*
- *How best to manage any transition between the current situation and the new one, including a revised regulatory apparatus;*
- *How best to monitor the success of the new arrangements and tune them as necessary.*

In principle a Public Data Corporation, with suitably amended Terms of Reference, might act as the fulcrum of making the new system work. It *could* be the custodian of the National Information Infrastructure. The Corporation could act as the umbrella under which Open Data could be promoted and facilitated along with key matters such as attention to user requirements¹⁷, standards, specifications, quality, review, inducements, regulation and guidance, monitoring operational efficiency of government providers and supporting both data providers and data users. It could act as the trusted arbiter in finding answers to the difficult questions. Operating as a centre of excellence, it could be of real benefit in the information economy and would help to promote efficiency, cross-fertilise good practice, reduce duplication, and to stimulate the considerable growth that appears possible. If set up appropriately, it could also foster the new legislation that appears necessary. We would wish to see the same principles applied across the whole UK¹⁸.

We could see various organisational structures working but the PDC would need to have some degree of independence to have credibility. For example if set up in this way it could advise government on the PSI implications of plans for privatisation (e.g. as in the case of the planned privatisation of the Royal Mail, including the Postcode Address File¹⁹) or outsourcing²⁰. As indicated earlier we are not opposed to some charging for data and information but are convinced that any public-owned organisation which legitimately trades in data, information and services in future should be constituted as a government-owned company, and regulated as such. As in the case of NHS hospitals this requires a failure regime to be constructed.

We close this section of our consultation response by re-iterating our support for the Transparency Agenda and Open Data. As indicated above, we also see a positive role for a Public Data Corporation but not the one we have deduced from official statements made to date. We stand ready to discuss with government the three fundamental 'unblocking' mechanisms identified above - sorting out a coherent PSI strategy for the public sector built on a National Information Infrastructure²¹, crafting holistic legislation to reform the current 'patchwork quilt' of statutes and regulations and finding a better long term funding solution given the unintended consequences of the Trading Fund model. We also believe that, if government data are to be charged for, it is critical that there is a suitable governance structure for deciding pricing and other terms. Without such steps we are convinced that – despite best intentions – little will be achieved.

¹⁷ Inside and outside government. We note that a precursor to this function has been set up to oversee and provide advice to government on the public sector's requirements for Geographic Information, and how it should be sourced <http://www.communities.gov.uk/news/corporate/1878453>

¹⁸ The differing geographical or jurisdictional remits of organisations like OS, Met Office and HMLR is a complicating factor.

¹⁹ Where this occurred in the Netherlands, we understand that the Dutch government subsequently bought back rights to the equivalent of PAF.

²⁰ As government explored for the Land Registry

²¹ Some of this can be launched immediately (e.g. the strategy) without waiting for the launch of new legislation.

More generally, APPSI members are deeply aware of the challenging circumstances in which policymakers find themselves in at present. But, in the absence of financial and other resources, a sound plan and good governance, aspiration and lofty rhetoric can easily exceed the ability of public servants and public bodies to deliver government's ambitions. APPSI is pleased to help in any way it can.

SECTION B: ANSWERS TO THE SPECIFIC CONSULTATION QUESTIONS

PLEASE NOTE THAT WE HAVE EXPRESSED MAJOR CONCERNS WITH THE CURRENT CONCEPT OF THE PDC IN THE PRECEEDING SECTION OF OUR RESPONSE. RESPONSES TO THE QUESTIONS BELOW REFLECT THESE CONCERNS.

Question 1. How do you think Government should best balance its objectives around increasing access to data and providing more freely available data for re-use year on year within the constraints of affordability? Please provide evidence to support your answer where possible.

APPSI Response:

- We urge taking a more strategic view than has been the case thus far, focusing on the National Information Infrastructure (Annex 1) and prioritising key data resources. Placing more data sets on www.data.gov.uk does not necessarily meet user needs better than improved access to the key data sets. Key data sets include definitive addresses and street networks (see also answer to question 2);
- All organisations must now become ever more efficient and we assume that requirement will continue. The temptation is therefore to extract more resources out of these organisations. Whilst such improvements are certainly possible it is much easier to save money by degrading the quality of data – collecting them less accurately, less frequently or via surrogates. We believe the sustainability of the National Information Infrastructure depends on maintaining sufficiently high quality databases (the products which can be spun off are less important since they can normally be re-created relatively cheaply) and this must be a priority;
- Currently there is little incentive to take risks in setting prices to grow a market and reduce unit charges over a number of years; public sector data traders' incentive is to maintain income from existing markets. If government wants to maximise PSI re-use it should seek a mechanism for at least sharing risk with those data providers.
- Since we can not tell what affordability will be in future, we have difficulty in giving more precise answers – other than needing some 'user champions' embedded in appropriate governance structures to ensure that key data bases are not allowed to degrade under financial pressures;
- By ensuring that the wider and longer term perspectives on value continue to be factored into funding decisions. It is very easy to discount any returns which are not manifested in terms of immediate cash but rather in longer term tax income from the private sector and better decision-making and citizen engagement;
- By understanding (and simplifying, as we have proposed) the complex and confusing sets of legal structures under which PSI access and re-use operates. The virtual absence of mention of the INSPIRE Directive and the EIR regulations from the consultation give us cause for concern that navigation of the legal hurdles will not be carried out effectively;
- Ensure that metadata is adequate for the 'normal user' thereby minimising the user frustration, enforced engagement with the data providers, potentially misleading use and additional costs;
- Recognise that changing technology, enhanced expertise and greater capacity in the private sector may offer a quite different way of operating data collection and dissemination at lower cost.

Question 2. Are there particular datasets or information that you believe would create particular economic or social benefits if they were available free for use and re-use? Who would these benefit and how? Please provide evidence to support your answer where possible.

APPSI Response:

Earlier we argued that “APPSI members are clear that access to ‘Core Reference Datasets’ – those where there is a moral right for the public to have access (e.g. details of all the UK laws) or where there are large safety, public service, efficiency and cost benefits if everyone uses the same definitive and regularly up-dated sources of data – should be available at no more than marginal cost. In an Internet world this translates to virtually free at the point of use.”

The following list is partial but marginal cost or free at the point of use, including ‘uncharged for re-use of the data sets listed below – some of which could certainly be considered Core Reference Datasets - would have considerable benefits:

- Addresses (Geoplace NAG);
- Addresses (Royal Mail PAF);
- Streets (Ordnance Survey ITN);
- Council tax band information for individual properties (VOA);
- Other Council Tax and Non-domestic Rates property attributes (VOA);
- The ONS Postcode Directory (the definitive look up table which allows any data with a postcode to be aggregated to a wide range of statistical and administrative geographies). Currently charged at only £200 this would obtain much wider use if made available free for downloading without restriction on re-use;
- Inter-departmental Business Register (ONS);
- More government statistics at smaller (Output Area) level;
- More anonymised unit record survey data (c.f. ONS Census Sample of Anonymised Records – SAR’s);
- Given the widespread use made of climate data and weather forecasts using (free to use) commercial web sites (e.g. for holiday planning) there must be some greater value that could be extracted from Met Office historical and contemporary data for the UK and beyond. Whether this will produce significant commercial returns is another matter given the international competition so we suggest it is made available as Open Data (see the definitions in our response to the Open Data Consultation);
- Land Registry data are used to produce various spin-offs like a House Price Index (one of several produced by the UK government). The cost of aggregation of these data from individual house purchase records can not be high. So if little revenue is generated we suggest this is simply made available free of charge.

Some of these would manifestly require a decision to lose trading fund income while some might need support to overcome real or perceived barriers from current legislation that underpins the data.

Evidence: The list is based on years of experience of some APPSI members working with organisations that have decided they cannot afford or are denied access (e.g. to SAR’s) to these priority data sources but would benefit greatly from free availability in making better decisions, improving research or making financial gains. Many are currently using inferior

products to save money. These include (perhaps surprisingly) various government departments, local authorities, academics and even some utilities and commercial companies.

Question 3. What do you think the impacts of the three options would be for you and/or other groups outlined above? Please provide evidence to support your answer where possible.

APPSI Response

See our more radical strategy outlined earlier which we think is a much better solution; in it we specifically comment on the funding (and hence charging) issue. APPSI members are clear that access to 'Core Reference Datasets' – those where there is a moral right for the public to have access (e.g. details of all the UK laws) or where there are large safety, public service, efficiency and cost benefits if everyone uses the same definitive and regularly up-dated sources of data – should be available at no more than marginal cost.

In an attempt to be helpful however, our comments on the three specific options cited are as follows:

Option 1: There is not much point in having a PDC as proposed if this key element of activity is to remain²² the same; we also query the meaningfulness of the statement "there *may* be a role for a regulatory oversight function in *advising* PDC and its constituent parts as to how they *might* go about making practical arrangements to make *more* freely available data year on year, *subject to affordability and value for money*" [our emphasis].

Option 2: There is an intrinsic attraction in harmonisation and simplification; there is no doubt that any user who wishes to obtain data from more than one of these different organisations would find the pathways to obtaining it significantly different and probably confusing. We do not however know how many such users exist or might exist.

The simple model proposed of charging the same for any use of the data and levying a charge per unit of data has two difficulties. The first is that this model – unless all charges are low – will foster high value applications and reduce ones with non-monetary (e.g. research or social) benefits. The second is that the concept of a unit of data is impossible to implement across a wide range of data sets. Moreover, it is easy to 'bulk up' data unit numbers or even the number of micro units (e.g. spread sheet cells, spatial co-ordinates). The proposed charging metric is therefore open to great abuse and is in practice likely to cause angst and accusations of unfairness.

One issue linked to charging which should be considered is derived data. There are many challenges associated with this, including re-use charges (e.g. when they should apply and how much) and there could well be a conflict with a standardised charging scheme. For example, an innovative product could seek to use data from several sources (which could be both inside and external to the PDC). If, for each source, there is a standard condition that any product which uses information and data under a chargeable licence has to pay 20% of the selling price to each originator (even if a very minor part of the whole), then the result could quickly become unprofitable. We note that Option 2 for charges appears to rule out a charge when derived data products are sold. Was this intended?

²² We see inconsistencies in current practice and how it is described in the consultation. For example, we note the statement in the Met Office pricing policy that "HM Treasury Guidance states that where the Met Office provides non-competed public services (including services to other public sector bodies), the standard approach is to price such business based upon full cost recovery plus a real charge for the cost of capital and may include an agreed efficiency target". Some of this will be part of the Met Office's public task.

Option 3: We see some attraction in Freemium mechanisms and price discovery systems are quite sensible, where pricing for information can sometimes be optimal by discriminating across users. Whether this is acceptable in a government environment is not clear. Based on experience in the transport sector, we also have reservations about the extent to which such approaches can be effective across the PSI landscape and are sustainable over time. The essence of many parts of our information infrastructure however is that they must be predictably and regularly up-dated and maintained”.

Question 4. A further variation of any of the options could be to encourage PDC and its constituent parts to make better use of the flexibility to develop commercial data products and services outside of their Public Task. What do you think the impacts of this might be?

APPSI Response:

We have earlier criticised the multiple roles envisaged for the PDC, saying “The consultation proposal conflates mutually incompatible functions together: operations, sales and marketing, standards setting, procurement, policy-making, and a vehicle for private investment “.

We believe this would cause continuing problems and conflicts with the private sector and, in particular, be wholly inconsistent with any policy role for the PDC. Some of this has already occurred in the last few years leading to fierce debate about a lack of level playing fields, derived data considerations, and complaints to competition authorities. It seems certain that these would be multiplied by this proposal.

We are equally uncomfortable with suggestions of attracting private sector investment under the PDC umbrella. We say this because of competition reasons: it is entirely likely that a government-based body would have some ‘insider’ advantages. Moreover, we pointed out earlier that “If the PDC was a vehicle that sought to attract private investment, it would either have to extract monopoly rent from particular datasets or go beyond the PDC’s public task to compete with existing value-added resellers. This seems unwise in a government-based organisation (as we understand it to be”.)

Even if a joint venture (JV) were to be used as a collaborative vehicle to create added value products, there could be pressure to help sustain the JV by allowing advance access to information, additional investment (either direct or hidden, e.g. in kind) or a different charging mechanism. To help to ensure a level playing field, any JV would need to be managed at arm’s length and the Directors of the JV should not be Directors of the PDC or any part of it. Copyright issues would require careful control; in particular there should be no shared copyright or the JV having any copyright or licensing control for any PDC information.

If private sector capital is to be sought the cleanest and least troublesome way is for a complete break to be made and set up a separate ‘downstream’ trading company – which could be owned as a GoCo. It would need to run as a commercial activity, have considerable transparency about its operations and sources of funds and operate in compliance with competition law, i.e. be unable to do exclusive or unduly favourable deals with an upstream ‘data provider’ entity. To gain credibility would be difficult if the PDC entity was to be awarded contracts for government work which were not competed. This is not simply a commercial view: the OFT argued in their CUPi report²³ for the need for proper upstream and downstream accounting separation of PSI which is to be sold commercially.

Fair competition is difficult to ensure in an entity with an integrated database in any case since different products are spun off from survey information collected at different times, for different initial purposes and by different means and sometimes by different players. Under these circumstances cost allocation and calculation of profitability is manifestly difficult. It follows that

²³ Office of Fair Trading *The Commercial Use of Public Information (CUPi)*, December 2006

a division of 'upstream' data collection, assembly, checking / quality assurance and integration in the database from some (probably multiple) 'downstream' value adding agents is likely to minimise the problems identified and foster competition.

Question 5. Are there any alternative options that might balance Government's objectives which are not covered here? Please provide details and evidence to support your response where possible.

APPSI Response:

Individual members have given thought to other, distinctive approaches to meeting government's objectives²⁴. Some are revolutionary in terms of privatisation or outsourcing of government functions. Another APPSI member has proposed the creation of a PDC with some similarities to the BBC's market panel and BT's Equality of Access Board; its members would be independent but probably chosen from the public and private sectors. They would decide what public sector data would be made available to third parties outside the public sector and the basis upon which the data would be made available.

Some aspects of these proposals are explicitly ruled out in the PDC consultation or by recent government actions (e.g. in setting up the Public Service Mapping Agreement). Moreover, these ideas have not been debated collectively by APPSI so we only report them briefly here but will offer further input if that would be helpful.

Question 6. To what extent do you agree that there should be greater consistency, clarity and simplicity in the licensing regime adopted by a PDC?

APPSI Response:

Depending somewhat on the form the PDC takes and its final role defined by government, it is critical that it has a strong, transparent and effective regulatory framework (see answer to Question 10).

We totally support this aim though noting that – thanks to The National Archives (OPSI) – the situation has improved considerably in recent years.

Evidence: See the great success achieved by changes to Population Census data access and licensing description described in the earlier section.

In the licensing options, there may be a need to separately consider how Public Task information and other PDC information are licensed. For Public Task information, Option 3 seems likely to get very complicated and would cause additional work to manage. APPSI members are divided on whether Option 1 or Option 2 is to be preferred. Option 2 offers simplicity and clarity, especially for users working across different data providers. We do not know how easy it will be to operationalise given the many different types of data in the PDC which could have very different characteristics (for example, for some purposes, the value/'half life' of a meteorological observation reduces quickly over a matter of hours; whereas, for other data types, the frequency of changes and hence some part of the value can be measured in days or years). In circumstances involving use of data from different sources Option 2 might need to include several clauses, even in an overarching agreement that could vary between different parts of the PDC. We recall that there are already binding international agreements on charges and re-licensing which will have some impact on the option chosen. We strongly urge that a pilot should be set up to create a prototype Option 2 model and test whether it works.

²⁴ See Barr R (2011) *A Public Data Corporation: What is it? Why do we need it? How should it work?* Paper given to Association of Geographic Information Annual Conference

Question 7. To what extent do you think each of the options set out would address those issues (or any others)? Please provide evidence to support your comments where possible.

APPSI Response:

Nil response other than that embedded in the above statements

Question 8. What do you think the advantages and disadvantages of each of the options would be? Please provide evidence to support your comments

APPSI Response:

Nil response other than that embedded in the above statements

Question 9. Will the benefits of changing the models from those in use across Government outweigh the impacts of taking out new or replacement licences?

APPSI Response:

Nil response other than that embedded in the above statements

Question 10. To what extent is the current regulatory environment appropriate to deliver the vision for a PDC?

APPSI Response:

Governance and regulation must have the full confidence of re-users and all other stakeholders (not just shareholders). Fair competition is the key issue: it will require an independent regulator with teeth, sanctions and a rapid and low cost adjudication process.

If, however, data are not going to be exploited commercially by government, then the issue becomes more one of adjudicating the limits of Public Task and a different – but relatively low cost - approach would be required which OPSI as part of The National Archives, might well be in a position to provide.

It has been pointed out in the Consultation document that few complaints have been received under the current regime. The commercial view is there may be a simple reason for this: making a formal complaint is expensive and time-consuming, the PSI Regulations are unclear, OPSI's (The National Archives') present power is limited and the OFT has no remit to take action against Public Undertakings. Anecdotal evidence suggests that users lack confidence in the process.

We believe that The National Archives – with more resources and a suitable statutory basis -is well suited to become the regulator in either of these circumstances. It has proved effective in its activities thus far. We see the Information Commissioner's role as more of a policeman and – whilst this is an important function – the regulator needs to understand government Open Data principles and aims and commercial issues as well as privacy, data protection and similar issues. And the regulator must be committed to maximising the availability of Open Data subject to relevant legislation.

Question 11. Are there any additional oversight activities needed to deliver the vision for a PDC and if so what are they?

APPSI Response:

The consultation states that all data made available free for re-use would be under the OGL. This does not address the potential problem when data are made available free for some purposes (e.g. non-Commercial Research or Educational) but not others (e.g. Commercial applications). This needs to be addressed.

Some European countries have tried to reduce conflicts of interest by adopting an approach to corporate governance through a Supervisory Board and a separate Operational Board. This model might, reflecting common business practice in mainland Europe, be relevant for the PDC as envisaged under the government's current model whereby HMLR, OS, and the Met Office would report to a PDC Supervisory Board. In general however we favour simplicity in governance to the greatest extent possible whilst meeting the overall needs.

We have argued the case for a National information Infrastructure at some length. Central to this is the needs of all governments in the UK – not just Whitehall's needs. It follows that a suitable governance model needs to exist which covers all the relevant governments and, indeed, the whole of the public sector.

This may present greater problems than simply dealing with central government. The experience of local authorities – who have put a lot of effort into tailoring information to meet the (usually researched and tested, but certainly experienced) needs of people in their area, particularly when it comes to accessing services and influencing decision-making – is sobering. They have been required to release a lot of the underlying data, but so far there is little evidence of private actors making much use of this (other than for passing attempts at identity fraud). This may be because the data released so far is indigestible, and/or uninteresting, or because – as we suspect – most people want their data 'ready-prepared' as information. This buttresses our conviction that prioritisation is essential and that the success of the Open Data agenda will be a long term (but, we believe, a highly valuable) one.

Question 12. What would be an appropriate timescale for reviewing a PDC or its constituent parts public task(s)?

APPSI Response:

For the regulatory oversight, we consider that the maximum period between reviews of the public task of the PDC or its components should be two (or, possibly, three) years, at least in the first instance. This would allow the currently rapid changes in understanding, technology and the overall trading environment to be considered suitably frequently.

**Advisory Panel on Public Sector Information
27 October 2011**

ANNEX 2 RE-SHAPING THE NATIONAL INFORMATION INFRASTRUCTURE²⁵

Background

Recent months have seen a transformation in government's thinking about the value of Public Sector Information (PSI) and actions impacting on how it should be made available. There has been a dramatic freeing up of access to some data previously requiring payment and intimidating licensing. And the www.data.gov web site has centralised and simplified access to thousands of data sets drawn from across government.

APPSI has long supported the case for such radical steps and strongly welcomes these developments.

Thus far however most of the developments have been piecemeal. There has been relatively little strategic thinking about the whole of the public sector information system – what principles should underlie what we should collect, what should be charged for and what are the interdependencies between the myriad data collected by and for all the different governments in the UK? Moreover, the current financial cut-backs are already leading to reductions of expenditure on data collection, etc. within individual government departments without any consideration of the functioning of the whole system.

The purpose of this APPSI discussion paper is to tease out the big strategic issues needing to be addressed and set out APPSI's views on how best to proceed. Our approach is to seek to establish principles but also to be pragmatic.

The National Information Infrastructure (NII)

Governments need information to:

- assess systemic risk to health, well-being or business growth and guide policy formulation
- enable the public to hold government to account ,
- enhance the quality of services provided or enabled by government and to
- foster the creation of new enterprises fuelled by PSI.

The existing collection of data and its conversion into information (and evidence) results in a *de facto* National Information Infrastructure. This is analogous and parallel to the National Infrastructure Plan²⁶ (NIP) launched by the Prime Minister on 25 October 2010. The NIP concentrates on physical infrastructure such as roads and rail though it recognises the importance of broadband 'pipes' for carrying all types of information; it does not address the content of the information flowing through these 'pipes'.

In all of what follows we do not consider matters of national security though much PSI is relevant to that (shortly after 9/11 the US government funded a study by the Rand Corporation which reviewed whether the widespread availability of public sector information facilitated the selection of targets by terrorists). With some 85% of the critical physical infrastructure in the USA and about 30% of it in the UK being in private sector ownership, information about it is both confidential and spans the two sectors.

²⁵ This document is a preliminary one more work is needed to meet the requirements recommended in the consultation response.

²⁶ http://www.hm-treasury.gov.uk/ppp_national_infrastructure_plan.htm

APPSI contentions about the present and future of the NII

We accept that the world of information has changed irrevocably: changes in technology have transformed our ability to assemble, check, disseminate and use information. There is - and will continue to be – a marked reduction in the state's ability to control what is done with the information it collates. It may also be that 'crowd sourcing' will come to play a significant role in data collection both by the state (e.g. by the Met Office in gathering reports of snow fall) and by the private sector. And the present trends are leading both to the creation of deepening natural monopolies in some government bodies and disintermediation of others by the private sector. So it is impossible to predict with any accuracy the evolution of a NII which has been founded on past technologies and policies and which is encumbered by historical legacies.

Nevertheless the crucial role that the NII plays in the running of the country and the public expenditures involved in maintaining it require us to seek areas where we can shape the NII to be more effective. We set out below some contentions based on both long experience and APPSI members' engagement in many relevant fields.

1. The existing NII is based on an assembly of data derived from many public and private sources, often with the data being re-processed by many players. The UK is one of the countries with the most decentralised (and hence disparate) forms of data collection. The existing NII is therefore a 'patchwork quilt' of data sets collected and processed to different standards, currency and form.
2. As a consequence of this and little previous focus on the whole system, we know almost nothing about the effectiveness and efficiency of this system even though billions of pounds are probably spent annually on maintaining it. Assessing the quality of evidence derived from putting together multiple sets of PSI is also often only possible on a highly qualitative basis.
3. Information differs from the physical infrastructure in some characteristics. It is normally indestructible but, though it tends to have a half life in its value, old information can be hugely valuable for trend analysis. Substitutability is sometimes easier (e.g. official statistics are sometimes substituted by those from private sector trade bodies) and data can often be created faster than, say, a new motorway (though it typically takes several years before a new data source is brought on-stream by government).
4. Standards for many aspects of physical infrastructure are well-defined (even internationally) but those for data collection by government departments are often a matter for internal decision by each department, sometimes after wider consultation.
5. The UK government has moved to a situation where it is increasingly the *de facto* monopoly supplier of foundation data (e.g. in subsidising the 'free data' now provided by Ordnance Survey and creating a pan-government agreement to supply OS' data to all of the public sector). The policy appears to be that most foundation data will be supplied at marginal or zero cost²⁷ and that the private sector will provide value added activities.
6. However the lack of any cohesive framework to define the public task of official data creators ensures that the private sector is nervous of making investment in this downstream market because it sees government bodies increasingly treating this as a revenue generating mechanism to replace reducing income voted by parliament. It is inevitable that changes in data and information requirements will occur over the coming years so there must be a process in place not only to define the public task of government bodies so far as data collection is concerned but also to re-assess it periodically. And it is critical that the definition process is not controlled by the data collectors or even simply their parent bodies.

²⁷ See Prime Minister's Open letter to government departments May 2010

7. The open data policy thus far has been posited on the proposal that ‘make it available and the users will come’. It is too early to see if this will work well beyond its enthusiastic welcome by technical experts and, if so, in what areas will the change be lasting and beneficial. But what is certain is that without the provision of metadata²⁸ and much greater levels of description the interpretations made will be much less accurate than should be the case. Already we see gross misinterpretations of government data daily in the media, by parliamentarians and others. We are not naive enough to think this can be reduced to zero. Robust debate and the low level of mathematical understanding in Britain²⁹ by the standards of some other countries ensure this is an impossibility. But we do believe that the provision of descriptive information about government data, information and evidence – such as how it has been compiled and by whom, the quality assurance processes used and the likely accuracy – is a crucial element in minimising misinterpretations.
8. The NII is no longer national. The EU INSPIRE Directive has major implications for UK data collection and the Environmental Information Directive impacted heavily on UK pricing policy in some areas. In addition, the activities of the private sector are having substantial impacts: data and tools provided by Google, for example, are now widely used in UK governments and by business and the populace. This international dimension raises important policy issues of equity (e.g. UK firms having to pay for some meteorological information from UK sources whilst US firms can obtain the same data free from US sources) and what we continue to need to collect ourselves.
9. There have been numerous changes in the organisational structures of government bodies collecting data in the last two decades (e.g. the Coal Authority, Network Rail) and others are in prospect - e.g. the Royal Mail and the putative Public Data Corporation). The result is a variety of ways in which the original PSI is made available, if at all. This variation in of approaches cannot be appropriate. If data are now collected by an agency on behalf of government, they should be subject to the same rules as if they were collected directly by government. If the data were originally created by a government body at public expense, then the successor bodies should be required to follow government policy on data dissemination, charging, etc.
10. Finally, as observed earlier, we see continuing shifts in the thinking on what data should be charged for and what should be supplied free. The ‘end of history’ view that acceptance had now been reached by all in UK government on this matter is premature.

APPSI’s views on the way ahead

These are that:

1. There is an urgent need to think more widely about government information than is normal. A speedy high-level review of the NII should be set up (and, since HMT has expressed an interest in data collection and interdependencies, it may be the best sponsor though we recognise the responsibilities of the Ministry of Justice for information policy and the interests of the Cabinet Office in such matters). The review needs to consider the likely impacts of globalisation on relevant aspects of data policy and collection.
2. There needs to be a standard process for consulting within and outside of government on proposed cutbacks to components of the NII (a number of government departments are already engaged in consultations). The results of these however need to be

²⁸ ‘data about data’

²⁹ http://www.pisa.oecd.org/pages/0,2987,en_32252351_32235731_1_1_1_1_1,00.html;
<http://www.nuffieldfoundation.org/>

aggregated, taking into account interdependencies, to assess the systemic effects of changes made under financial pressures).

3. Where data have been collected by or for government it should continue to be regarded as PSI, irrespective of changes of structure of the relevant agency.
4. Renewed discussion and consultation on the 'rules of the game' in terms of charging (and, if so, how) for data and information is essential. It is simply unrealistic for on-going data collection and free dissemination to be continued (or expanded, as some suggest) when expenditure reductions of some government departments by 40% and local government by around 25% are taking place.