

APPSI RESPONSE TO THE OPEN DATA CONSULTATION

INTRODUCTION

This statement forms a response to the publication of a public consultation: ***Making Open Data Real***. A separate response is provided to the partner consultation: *Data Policy for a Public Data Corporation*.¹ We begin with a high level contextual statement designed to identify and address some big issues for the forthcoming White Paper on Open Data. This is followed by specific answers to the direct questions posed in the consultation paper.

APPSI is an independent body² 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 Process 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 areas. 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 OPEN DATA AND THE CONSULTATION

We reiterate our strong support for the agenda to make more widely available those data created by or for governments in the UK (PSI). We are convinced that this will lead at various levels to the benefits postulated in Section 7 of the Open Data consultation. And, recognising that collecting data has (often significant) costs, we note that many organisations are not in principle opposed to paying for some data - although we would expect some key data sets to be free at the point of use and that charges should be both fair and should be reduced over time as better information management practices develop and markets grow.

The Open Data consultation is a thought-provoking and serious attempt to tackle a complex and genuinely multi-faceted set of issues. We welcome its contribution and many aspects of it, notably the need to take a more strategic approach. We address the complications below.

¹ Government has also recently published a third document which bears upon PSI availability: *The Transparent Government, Not Transparent Citizens: A Report on Privacy and Transparency for the Cabinet Office* was published on the day of our last APPSI meeting. This has potential implications for Open Data and we will respond to it appropriately.

² <http://www.nationalarchives.gov.uk/appsi/default.htm>

HIGH LEVEL RESPONSE: PSI IN THE UK

We agree with the 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”. We note that Tim Kelsey, Director of Transparency in the Cabinet Office, is reported³ to have argued, “...lasting transparency will not come from an episodic approach to releasing data: we must develop a more strategic approach”. APPSI has long argued for just such an approach. This statement should therefore be viewed as a constructive attempt to support such change.

Addressing the complications in re-use of PSI

Terminology. The variations in language used by the many different players foster misunderstanding, lack of clarity and confusion. This lack of clarity extends to the two consultations. Our own suggestions about a more consistent terminology are set out in Annex 1.

Custodians and beneficiaries. We take it that government bodies which collect or assemble data are not owners but rather are custodians of that data. The data, as we understand it, has four beneficial purposes:

- To enable government to analyse the current situation, make policies and monitor their progress in achieving their aims;
- To help government to deliver its services most effectively;
- To provide a public good so that members of the public can monitor the performance of government and the probity of those in authority, notably in delivering services to them, and can understand the state of the country or their local community;
- To support innovation – especially through the private sector - and, through it, to create new jobs and a better quality of living;

Open Data policies need to address these four sets of beneficiaries.

Managing expectations. Open Data is necessarily something of an experiment since the UK is one of the leading nations in this area and the underlying technology is developing rapidly.

Few if any sets of Public Sector Information will support all of the six benefits⁴ identified in the Open Data consultation. For example, data on government’s performance against targets and senior civil service staff salary data may well support the Transparency Agenda but they seem to us to have limited obvious commercial value. Some data sets such as crime statistics and maps made from them are highly appreciated by the public but may not be possible to monetise. On the other hand some real time data on the status of transport and traffic plus some ‘Core Reference Datasets’⁵ seem to have significant commercial or wider economic value. For many others of the 7000 data sets now available on www.data.gov.uk the actual benefits they offer are wholly unknown at present.

³ Civil Service World 5 October 2011, page 3

⁴ See Section 7 of Making Open Data Real, pp19-21

⁵ Though we argue that these should be free at the point of use (see APPSI’s response to the Public Data Corporation consultation)

It is manifestly the case that not all data are of equal value. To date however the Open Data consultation seems to be working on this premise. Prioritisation of 'suitable cases for treatment' is essential if expectations are to be fulfilled⁶.

We understand the merits of taking urgent action taken by government but we believe that the fundamental changes which are being sought in behaviour in opening up government data to scrutiny and for re-use will only succeed over an extended period and with a clear strategy.

'Free' data or Charging? We have dealt with the issue of charging at length in APPSI's response to the Public Data Corporation Consultation. Here we simply make a brief summary.

There are costs, sometimes very substantial, in collecting high quality data; who pays – taxpayer, user, the originator of change – is an important consideration. But in his foreword to the consultation, Francis Maude MP observes that "the best way to tap into the UK's tradition of creativity and invention is to give that [PSI] data away". This presumption of publication of data for free re-use is now a government policy principle though it is heavily caveated in the two consultation documents.⁷ There is already a widespread recognition that some data sets should *definitely* be made available free of charge to end users. The legislation database is one such example, based on the principle that all citizens should have equal access to the laws by which they are governed. Official statistics fall in to the same category. We would expect that Core Reference Datasets would be made available free at the point of use and that all 'exhaust data' – that produced as a by product of government activities for internal purposes – would also 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.

The evidence for best practice over charging is contentious. Free at the point of use certainly increases take-up, sometimes by orders of magnitude⁸. Notwithstanding various heroic studies, there is no general agreement on whether greater economic benefits arise from charging for government data (attracting immediate revenue from licensing) or from fostering commercial organisations to innovate and market new products based on the data (and hence generate some downstream tax income). For many data sets this is because they are an 'experience good' i.e. their value is unknown until experiments have been carried out. In an ideal world we think productive use of PSI would hugely increase if it was all free at the point of use. But we recognise current financial circumstances and legacy agreements (notably the existence of Trading Funds created at different times and under differing imperatives). In our Public Data Corporation consultation response we argue that the Trading Fund is inherently inimical to the aims of Open Data and generates unhelpful behaviours in both the public and private sectors. When data are being considered for free at the point of use or inexpensive release, it is vital that the implications for the supply organisation are well considered. They will usually have some costs that have to be met and

⁶ We recognise that cost-benefit analysis of different types of data is very difficult (since data/information is often an 'experience good') but judgements need to be made

⁷ "to provide more freely available data for re-use year on year within the constraints of affordability"

⁸ For example the 2001 Population Census results which were available for free to commercial as well as public sector users, plus Ordnance Survey OpenData and the results of the Public Sector Mapping Agreement (PSMA). The latter has been taken up with some enthusiasm by government organisations and emergency services to improve their research, policy development, operational response and public services. The data was paid for, but centrally by government. Unfortunately, many organisations outside the public sector are not readily able to arrange such large scale efficient group purchasing (though some commercial organisations choose to negotiate prices on a one-to-one basis so as to preserve confidentiality of their planned use of data for competitive reasons). It *may* be that this is a role for a Public Data Corporation.

it is necessary to plan for enduring maintenance regimes that will uphold the quality of the data.

Quite a lot of evidence exists that at least parts of the private sector do not object to paying a reasonable amount⁹ for certain government data – provided it will enable them to generate efficiencies or to build a successful business. Their main concern is that it often takes significant amounts of time to conclude access to data, notably since interacting with ‘customer facing’ parts of government to understand the characteristics of data, etc. is often tortuous and – notwithstanding substantial improvements – licensing of data can often prove complex and costly.

We understand that the public sector operates under more constraints in charging for data than does the private sector. Whilst it is well-established public sector practice to charge differentially on the basis of use type and numbers of users, charging the maximum that the market will bear on an individual customer basis is forbidden. The only solution is to spin off different but related products from a database and have a suite of different prices. We believe it would be helpful to have clear and simple statements on such pricing. Notably, we feel that the concept of ‘marginal pricing’ can mean different things in practice and its application should be clarified.

If government data is to be charged for, it is critical that there is a suitable governance structure for deciding pricing and other terms. In part this already exists in that business cases for charging must already be approved by The National Archives. But we read into the Public Data Corporation consultation that this may be a task for that proposed new body. This governance framework must be clarified.

As indicated earlier the APPSI view is that a number of datasets should definitely be free at the point of use or very inexpensive – notably the ‘Core Reference Datasets’. These include geographical frameworks, notably a national address database which acts as a tool to link other data sets together, typically through location. Again the principle is that there are substantial safety, public service, efficiency, and cost benefits if everyone uses the same definitive and regularly up-dated sources of Core Reference data. An irony is that some of these are currently ‘charged for’, at prices outside the reach of some latent users. It is of vital importance to recognise that some form of continued government funding in – and ownership of - these core databases will be needed.

What is needed? The exact data which contemporary governments need to discharge their functions and support the public good is not known. We suspect that more information is collected than is needed but there may well also be crucial lacunae as well. Government needs an effective way of discriminating between the essential, desirable and irrelevant. We have already made suggestions about how to take this forward in a strategic way via a National Information Infrastructure, outlined in Annex 2¹⁰.

What data do governments hold? We do not generally know in any detail what data governments presently hold, let alone what form these are in or their accuracy or currency¹¹.

⁹ 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).

¹⁰ For which a consciously conceived National Information Infrastructure should exist (see our response to the PDC consultation, and also Annex 2). See the report “Value of Geospatial Information”, sponsored by the Local Government Association, for an example of evidence-based data specification. This used an economic modelling approach to produce a “Top 10” list of applications using geographic information that had the greatest potential within local government to make more efficient the discharge of their functions.

¹¹ We are not convinced of urgency for the pan-government plans to create comprehensive data inventories across all of government– this is at least the third occasion this has been attempted; progress thus far has been very patchy for the task is considerable if done well and user needs for discovery metadata are not well understood.

Publish quickly or well? There is a trend to putting out much data now rather than making sure it is fit for purpose. On balance, we tend to the view that early publication is desirable and that data should not be nursed and perfected before publication. But it is vital, with such an approach, that users always receive guidance (through metadata) to help them to appreciate the value and suitability of each dataset at its current state.

What is government for? There is no real clarity about the respective roles the public sector and the private sector should play in regard to the supply chain for PSI in the UK. Some government bodies are (at least) quasi monopolies who market their own products, some work closely with private sector partners and in a few areas the private sector has established a business based on PSI from which the public sector derives little immediate financial benefit¹². Complications that arise from this include the ability of the public sector body to set 'fair' price levels in the absence of competition, for that to be perceived as fair¹³ and the difficulty of taking risks in building a market (see below). Moreover, government is increasingly adding value by 'co-mingling' its own data with that from private sector sources, muddying the intellectual property rights involved. Given all this it is not surprising that there are strong views held in the private sector that, where government acts commercially, distortions occur and the result is sub-optimal for the citizen as well as the public and private sector.

Where public sector bodies collecting data are then trading in or selling data (especially where they have first mover advantage or statutory protection and are using public funds), it follows that careful regulation is required. This and the appropriate form of licensing in such circumstances is a vexed question¹⁴ addressed in our detailed responses later.

The traditional view was that government was simply a provider of PSI that it happened to hold. In some government bodies this has long since mutated into the paid-for provision of services and products based on information they spin off from databases, increasingly 'co-mingled' with data held by others. Sometimes this exists in alliance with selected private sector bodies and sometimes in competition with them. Recent thinking has migrated to seeing government as a platform¹⁵ rather than a producer of products (but see also http://blogs.gartner.com/andrea_dimaio/2009/09/08/why-government-is-not-a-platform/). 'Platform organisations' like Google and Apple characteristically embrace open standards; build simple systems that can evolve; develop for participation; learn from their users, and especially those who do unexpected things; lower barriers to experimentation; build a culture of measurement; celebrate developers; and do not 'reinvent the wheel'. This all has implications for the roles of the private and public sectors and needs to be factored into the forthcoming White Paper.

Finally in this regard, we see government still acting as a series of separate entities. We detect a disjunction between immediate and longer term incentives. Those (typically the Trading Funds) with a quasi monopoly position in some of their activities who charge for data are required to achieve five year plans and produce an annual Return on Capital Employed. They have little incentive to take risks in setting prices to grow a market over the longer term; their 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 to transition to a larger market.

¹² The Registry Trust Ltd and Dr Foster

¹³ For example, charges for the new National Address Gazetteer arising from the Geoplace consortium owned by Ordnance Survey and local government payable by any organisation or person outside the public sector range from about £25K p.a. for one user to £190K p.a. for an enterprise licence. That said, these prices are for national coverage; most commercial users require data only for a restricted geographical area and the charges reflect that.

¹⁴ However the licensing situation has been immensely enhanced in the UK by virtue of the Open Government Licence; The National Archive's Office of Public Sector Information has attracted much praise nationally and internationally for its work on licensing.

¹⁵ <http://techcrunch.com/2009/09/04/gov-20-its-all-about-the-platform/>

Who is in charge? Many different authorities have a policy hand in what has emerged, including the Cabinet Office, The National Archives, the Ministry of Justice, the Department of Health, the Shareholder Executive and BIS, and the devolved administrations. Others with influence include the Information Commissioner's Office, the Office for National Statistics, and the NHS. Those organisations answering to Parliament (such as the UK Statistics Authority) are subject to a different form of direction.

All this has resulted in some confusion. It has also had some significant practical effects (see 'The world outside of Whitehall' below).

In some cases the allocation of certain responsibilities is obvious. As referred to in our response to question 5 of the PDC consultation, there is obvious scope for OPSI as part of The National Archives, and the OFT to take a lead in guiding the principles on encouraging market development and compliance with Competition Law. Who is best placed to lead on Open Data policy more generally, including ensuring the full confidence of all stakeholders in governance and regulation – especially given the long term nature of the agenda - is rather less clear to APPSI members. We look to the forthcoming White Paper on PSI to provide an unambiguous statement of policy and responsibility.

Changing technology. Without recent technological developments the Transparency Agenda and the PSI re-use agenda would have been stillborn. We have observed a trend for collection of government data to change from being primary (e.g. through surveys of various kinds) to being based on pre-existing administrative data. Typically this is provided by individuals seeking benefits or services of one kind or another. The approach has the advantage of reducing burden on the state and providing more up-to-date data. But it has certain disadvantages: privacy is potentially an issue, the accuracy may be lower and variable and changes to the primary need for the data may result in the end of comparable time series, affecting the safe uses to which the data may be put. We think this is an issue which has not yet been widely enough addressed.

The world outside of Whitehall. The main focus in these two consultations has been on data from central government. In reality, many public bodies in the UK collect and disseminate important data (e.g. the devolved administrations, the NHS and Local Government) and their incentives for so doing vary widely. Available staff skills and finance are often severe restraints. Incentivisation of such organisations needs more attention.

In addition, the current situation is leading to obvious areas of difference between different parts of the UK: these include health or any aspect at all of local service delivery. Other differences between the administrations occur in such areas as flood risk boundaries, where different approaches to the risk analysis and the release of information generate discontinuities at the borders. Another example is the different policies in relation to the release of open geographical datasets between Great Britain and Northern Ireland. This in turn inhibits the creation of a genuine UK-wide picture which we hold to be important for a number of purposes.

International factors. Open Data policy is clearly of great interest around the world. Major international players (e.g. Google, Navtech, Microsoft, South Street and various meteorological businesses) have entered the data supply market and in some cases compete with British government and commercial data suppliers.

Finally, the EU plays a role through its Re-use Directive and the UK Regulations created to implement the directive. At an earlier stage the European Parliament's debates were focused on a much more radical version of the Re-Use Directive, in which the cost of reproduction and dissemination only (and not the costs of collection and production, let alone a reasonable return on investment) could be recovered - a view much closer to the definition of marginal cost. The difficulty was that because the directive proposed a 'one size fits all' approach it was not possible to get agreement on those cases where this sort of marginal cost based pricing might be appropriate, and those where a different charging model should be tolerated. The situation may however have moved on. Neelie Kroes, Vice-President of the

European Commission responsible for the digital agenda, stated on 22 September 2011 “...And so at the end of November, I will be proposing to my fellow Commissioners that we adopt our next steps on the re-use of public sector information, and a proposal for an improved Directive. I want requirements to be more encompassing, and specifications improved. In particular, we'll be looking at the way data is disclosed - the formats and the way data licenses operate to make re-use straightforward in practice. We'll also be looking at charging regimes because expensive data isn't "open data". In short, getting out the data under reasonable conditions should be a routine part of the business of public administrations”.

These international developments will impact on the UK's plans and will need to be factored into the forthcoming White Paper.

SECTION B RESPONSES TO SPECIFIC CONSULTATION QUESTIONS

An enhanced right to data	
How would we establish a stronger presumption in favour of publication than that which currently exists?	We are attracted to all three in the suggestions list. Whether a new independent body is required (suggestion 3) is not clear. The National Archives has a good record in regulation to date but will need statutory support to achieve the objectives we propose in our Public Data Corporation response in relation to regulation and governance
Is providing an independent body, such as the Information Commissioner, with enhanced powers and scope the most effective option for safeguarding a right to access and a right to data	See above. We see the Information Commissioner as a ‘policeman’ and believe we need a body which will actively encourage and support use within the (reformed) law – The National Archives is the obvious body. The Information Commissioner’s remit does not extend to Scotland. Other regulatory matters are however important, notably in regard to competition law and policy e.g. Government has, in the past, appeared to believe that existing regulation (IFTS, PSI Regulations etc) in this space is adequate, and the OFT has little or no power of sanction over Government Undertakings.
Are existing safeguards to protect personal data and privacy measures adequate to regulate the Open Data agenda?	Yes – if properly applied. Linkage of individual data within a secure government environment, leading to new aggregate anonymised public data, needs to be carefully planned and covered by appropriate safeguards (e.g. criminal sanctions for disclosure as apply to census data)
What might the resource implications of an enhanced right to data be for those bodies within its scope? How do we ensure that any additional burden is proportionate to this aim?	If the rules are drawn appropriately, the incentives would probably ensure that routine pro-active publication would minimise the burden on public sector bodies of responding to individual queries. But an experiment or trial would be needed to ensure the remaining burden was not disproportionate. However our overall view is that as greater efficiency

	will continue to be sought from all organisations we assume that the requirement for data access will continue. The temptation will be 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. It will be important therefore to ensure high quality databases are maintained within a finite but diminishing amount of resource.
How will we ensure that Open Data standards are embedded in new ICT contracts?	government has found it difficult to agree and then implement common standards in the past for reasons which are unclear but may relate to inadequate pressure being applied from politicians and department heads. Having common Open Data standards is essential to the success of the policy. An appropriate Head of ICT Profession should be held to account for such compliance. Failure to comply should result in financial penalties for the organisation or reports to Parliament.

Setting transparency standards	
What is the best way to achieve compliance on high and common standards to allow usability and interoperability?	<p>All three of the bullet point suggestions on page 26 have merit and are not mutually exclusive. In aggregate they will not add a significant burden to individual bodies. But some mechanism for challenge when departments fail to perform to these standards is required (alongside a 'bouquet' system for recognition of those who do perform well).</p> <p>We would wish to see consideration of applicable standards such as the European Union's INSPIRE regulations be made within the proposed policy on Open Data – there is little recognition of this in the consultation.</p> <p>Some of the data produced by parts of the public sector is subject to INSPIRE and its associated legal requirements. The consultation is focused upon the UK: some parts of the public sector work as part of a global data exchange network (e.g. for weather forecasting); and as such are also subject to the standards agreed as part of that work. Clearly having to support multiple standards to meet varying legal requirements could be very expensive and act as a strong barrier to fuller adoption of selected multiple standards unless there is some recognition of other requirements.</p>
Is there a role for government to establish consistent standards for collecting user experience across public services?	Yes but this must not be heavy-handed (and see above). The aim should be to achieve it by definition of what is expected and promulgation of the policy then use 'dip-stick' testing or through a complaints process.
Should we consider a scheme for accreditation of information intermediaries, and if so how might that best work?	This is attractive in that mis-use (often inadvertent through lack of understanding of the data or poor documentation) can cause havoc. This is not just a fault of intermediaries but it is an area where the

	failures should be dealt with.
Corporate and personal responsibility	
How would we ensure that public service providers in their day to day decision-making honour a commitment to Open Data, while respecting privacy and security considerations?	The key way forward is by placing a suitable public responsibility on the organisation's governing Board. External exhortation and even threats by a revised Transparency Board and sectoral equivalents will be less effective
What could personal responsibility at Board-level do to ensure the right to data is being met include? Should the same person be responsible for ensuring that personal data is properly protected and that privacy issues are met?	We see no difficulty about combining these two 'protective' roles but what is essential is that there must be a positive responsibility on some Board-level individual to ensure that Open Data principles and policies are followed through.
Would we need to have a sanctions framework to enforce a right to data?	We take the view that complex license agreements deter many potential users, who simply walk away, rather than face the prospect of a long dialogue with company lawyers. We very much welcome current government initiatives in this area and public servants "honouring the commitment to open data". Board ownerships, and KPI objective setting, were all to be commended. But we believe that efficacy will depend primarily on; <ul style="list-style-type: none"> (a) clarity of policy. (b) appropriate off-setting of marginal costs of provision (if a public body so chooses). (c) defining realistic long term expectations. (d) suitable incentives.
What other sectors would benefit from having a dedicated Sector Transparency Board?	Environmental. Should this not extend over all the government sectors embraced by departments and Select Committees?
Meaningful Open Data	
How should public services make use of data inventories? What is the optimal way to develop and operate this?	We agree that "a right to data is meaningless without knowledge of what is available" (page 31 of consultation). Yet the history of government data inventories – especially for those bodies producing only 'exhaust data' is poor. Attempts to create signpost or discovery metadata sites have typically failed to meet the needs of users, many of whom will already know what they want and have experienced other barriers than discovery. To be useful, data inventories need to be designed to suit the data available and knowledge of user needs. Detailed national guidance is therefore unlikely to be useful. Some experience of these complexities already exists e.g. the national statistics hub (where data for some of the devolved administrations do not appear); careful consideration of this experience must be built into any new system. The role of commercial search engines may in some

	cases be the best solution. Prioritisation of which inventories and metadata to create is essential.
How should data be prioritised for inclusion in an inventory? How is value to be established?	<p>See our contention in the first section of this response that a National Information Infrastructure be defined based on the best available evidence. Serendipity or even short term user requests are not strategic enough approaches to prioritisation. Priorities will also depend on objectives, e.g. transparency vs. economic value.</p> <p>We very much believe that Core Reference Datasets are priorities and comprise an area where there will be significant social value. Core Reference Datasets' are 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 (see more detail in our PDC consultation response).</p> <p>Earlier we argued that APPSI members are clear that access to these should be available at no more than marginal cost. In an Internet world this translates to virtually free at the point of use</p>
In what areas would you expect government to collect and publish data routinely?	Where it is needed for its own purposes and for public use in holding government(s) to account and for understanding of their communities, economic factors etc. Government should not prioritise data collected primarily to meet the needs of the private sector.
What data is collected "unnecessarily"? How should these datasets be identified? Should collection be stopped?	This can not be answered without a proper study as part of defining the National Information Infrastructure. Note that some insights into this should have been achieved through the reviews by departments of their statistical data collection now underway and the HMT review of data collected. Where proposals for curtailment of statistical data collection have been promulgated these have been followed by a public consultation. No such public engagement (or even description of findings) has yet been published in regard to the HMT study.
Should the data that government releases always be of the higher quality? How do we define quality? To what extent should public services 'polish' the data they publish if at all?	<p>Data quality currently varies considerably. In principle what is collected should be designed to be 'fit for purpose' and no more. This of course demands knowledge of the user needs. It also demands a sound methodology for describing quality (not easy) so that latent users might assess whether the data are suitable for their needs. Methods of collection or provenance should be described as well as any known 'warnings'. We suggest that a quality metadata template be applied to all data but this may need to differ for different kinds of data (see above.).</p> <p>Making a trade-off between data quality and speed of publication is inescapable for many economic and some other data sets. Again the solution lies in understanding user needs. A code of practice may be</p>

	<p>helpful in defining good practice e.g. the official statistics code requires that data publication dates must be pre-announced and as soon as possible after the data are ready.</p> <p>‘Polishing’ is a pejorative term: manifestly unnecessary enhancement of data should not occur. The problem is that in practice the distinction between ‘added value’ and ‘raw’ data is rarely clear: it again depends on the nature of the particular kinds of data.</p>
--	---

Government sets the example

How should government approach the release of existing data for policy and research purposes: should this be held in a central portal or held on departmental portals?	<p>Either suggestion could work (and redundancy may even be helpful e.g. the same statistical data accessible via www.data.gov and on departmental portals). The success of direct.gov is partly about a wide range of services (not data) available in one source and partly because it is easy to use. Departmental sites often relate well to their users’ wider interests.</p>
What factors should inform prioritisation of datasets for publication, at national, local or sector level?	<p>This question is impossible to answer meaningfully in abstract – it depends on the most important uses. Census data is used for different purposes at different levels of geographical aggregation. Departments and users should decide.</p>
Which is more important: for government to prioritise publishing a broader set of data, or existing data at a more detailed level?	<p>This question is impossible to answer meaningfully in abstract (and it ignores the question of temporal frequency as well as geographical resolution). We envisage that insights on this would be produced in the course of definition of a National Information Infrastructure.</p>

Innovation with Open Data

Is there a role for government to stimulate innovation in the use of Open Data? If so, what is the best way to achieve this?	<p>We note that central government was the catalyst which ensured that Ordnance Survey and the Local Government Association finally agreed to forge a definitive National Address Gazetteer after at least five years of wrangling. So government <i>can</i> play a crucial role and we are of the view that there is a role for government to stimulate innovation of use of Open Data. However government needs to be more strategic in prioritising its actions (see above), wary of making the selection and assumptions of private sector interest and not to presume the developed benefit at a time when what will be successful is far from clear.</p> <p>How this is best done relates to what we discussed in the section ‘What is government for?’ in our introductory section. If government believes that innovation can be fostered effectively within the public sector that leads to one conclusion. If, as a number of APPSI members believe, the private sector and research bodies are vital agents of innovation, then government can stimulate that innovation through innovative procurement in extreme form this model would entail defining suitable public tasks for the relevant public bodies and abjuring any role for them to</p>
--	--

	<p>engage in value-added activities or even outsource the running of operations to reduce running costs (along the lines of the model of HMSO divestment of the 1990s). We note of course that not all such schemes have cut costs or maintained standards.</p> <p>In general Government can stimulate by setting standards, by investing in research, and by encouragement partnerships</p> <p>In our response to the PDC consultation we have argued for a different role for the putative PDC which relates closely to this question.</p>
--	--

Advisory Panel on Public Sector Information

27 October 2011

ANNEX 1 A PROPOSED TERMINOLOGY

Discussions in the Open Data community are bedevilled by uses of what appear to be synonyms but which are invested with subtly different meanings (notably in the PDC consultation). We very much believe that harmonising and clarifying the terminology used within the 'Open Data' world will be helpful. The following is not the definitive terminology to be used, but offers a starting point based on a study of words used in the two consultations and upon discussion at APPSI:

In addition to the use of new terms in different ways within the Open Data community, we have observed much loose use of well-established terminology such as 'marginal costs'. We have not repeated the standard definitions below.

Definitions of terms

Public Sector Information or Public Data (used synonymously though they have some different connotations). We prefer Public Sector Information.

Open Data (as described in various parts of the Open Data consultation):

- not charged for
- available also for re-use without charge
- readily accessible and available under the Open Government Licence
- includes all data relating to public service provision
- includes all data currently available for no charge

Controlled Access Data

- Data which are *charged for* under some appropriate licensing model under the UK Government Licensing Framework. The mechanism for deciding that charging can take place and the basis for it (e.g. in a complex situation of spinning off products from a widely used database) is to be based on a business case approved by The National Archives (including all Trading Funds).
- Anonymised information about individual's e.g. longitudinal databases, Samples of Anonymised Records (SAR). These are distinct from 'hiding' all information about individuals, for example by aggregating them in totals for a given geographical

area or class. Access to anonymised individual data to *bona fide* researchers and others requires special arrangements e.g. ONS virtual labs and as in the Scottish Longitudinal Survey.

Closed Data:

- Data which are not made available to the general public for reasons of privacy, confidentiality, national security.

Types of data

There are many types of data. The table below sets out some in common parlance. Some of the examples could be placed in other boxes. The point of this is that it is a varied and complex field and not always easy to generalise.

<i>Data type</i>	<i>Data type</i>	<i>Data type</i>
Accountability e.g. Departmental spending, organograms – public administration generally	Environmental data e.g. water quality, beach quality, carbon emissions	Database compositions Assembled to support public work, e.g. energy studies, land uses, traffic counts
Transparency e.g. MP's expenses, registers of interests	Land and property data e.g. transactions, value, ownership	Data assembled for specific projects Re-usable results from studies, government, local authorities, etc.
Data created in public task A wide range of these, related to formal duties, but not all PSI is public task	Geological data e.g. BGS, mine records	Interview surveys Commissioned surveys, regular or ad hoc. Often private but can be anonymised
Performance statistics e.g. school results, health admissions and outcomes	Operational data Result of processes or regular observations, often using machine recording	Modelled data Many forms of models output useful data, often preferable to observed/sample data
Information collected from/about citizens Much is subject to privacy rules, but not all	Transaction data Financial or bookings – usually private	PSI modified and re-published by private sector e.g. ACORN, property values, historic mapping extracts Note – there are many other relevant datasets from the private sector with no PSI content
Information collected from/about businesses Much is subject to confidentiality rules, but not all	Science data Research findings, formulae, etc	
Core foundation data Spatial – e.g. addresses, streets Statistical – e.g. Census	Social and demographic data Mainly from surveys but also lifestyle data, online activity, etc.	

It is important to understand the uses to which government data are (and will be) put to assess what should be collected/provided. To make sense of this we need to classify the data types described above in terms of type of use. One simple classification is as follows:

Sample uses of data*

<i>Class of use</i>	<i>Characteristics</i>	<i>Examples</i>	<i>Benefits</i>
Contextual	Defines states and trends	Social and economic statistics e.g. how many people in different age groups in each area of country. Met Office forecasts	Greater societal understanding of changes; Planning of possible changes to pensions and retirement age
Integrating	Facilitates linkage of data by geographic area or by individual	Address, postcode (Geoplace, Royal Mail) data and OS geodesy and mapping as frameworks. Outputs enabled include crime maps	Greater public and government understanding of incidence of problems
Monitoring	Enables public assessment of performance by governments against pre-defined targets	Numerous e.g. in health	Improved and focused public services (with some risk of creating perverse incentives)

Note:

- 'Transparency' as presently enunciated is mostly legitimated on the third row above. But many interdependencies exist e.g. you can't have crime maps without the capacity to link reported crime to geography
- Some official data may serve multiple purposes e.g. economic statistics

ANNEX 2 RE-SHAPING THE NATIONAL INFORMATION INFRASTRUCTURE¹⁶

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 Plan¹⁷ (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.

¹⁶ This document is a preliminary one more work is needed to meet the requirements recommended in the consultation response.

¹⁷ 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 cost¹⁸ 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

¹⁸ See Prime Minister's Open letter to government departments May 2010

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.

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 metadata¹⁹ 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 Britain²⁰ 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.

¹⁹ 'data about data'

²⁰ http://www.pisa.oecd.org/pages/0,2987,en_32252351_32235731_1_1_1_1_1,00.html;
<http://www.nuffieldfoundation.org/>

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 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.