

## **“Making Open Data Real” Consultation: Response from HM Revenue & Customs**

### **Introduction**

HM Revenue & Customs (HMRC) welcomes the chance to comment on the broader elements of the transparency strategy. It is clear that transparency has come a very long way in a short space of time – and as the consultation document sets out, the potential benefits of a well-designed transparency strategy are very powerful.

While we face some interesting challenges in respect of the interactions between transparency and our statutory duty of confidentiality, we are fully committed to the Transparency Agenda. We hope that our input and experiences are useful in shaping its next phase.

### **Contents**

| Section                                                   | Page |
|-----------------------------------------------------------|------|
| <a href="#">1 - Glossary of key terms</a>                 | 2    |
| <a href="#">2 - An enhanced right to data</a>             | 6    |
| <a href="#">3 - Setting open data standards</a>           | 14   |
| <a href="#">4 - Corporate and personal responsibility</a> | 18   |
| <a href="#">5 - Meaningful open data</a>                  | 20   |
| <a href="#">6 - Government sets the example</a>           | 24   |
| <a href="#">7 - Innovation with open data</a>             | 27   |

### **Contact:**

Kevin Reid  
020 7438 6023

e-mail: [kevin.reid@hmrc.gsi.gov.uk](mailto:kevin.reid@hmrc.gsi.gov.uk)

HM Revenue and Customs  
Knowledge Analysis & Intelligence – Project Group  
4<sup>th</sup> Floor South, Bush House, Strand, London. WC2B 4RD

## Specific questions posed in the consultation

### 1) Glossary of key terms

#### *a) Do the definitions of the key terms go far enough or too far?*

As specific transparency-focused definitions for the purposes of the consultation, the definitions work well enough at face value – particularly for the broader concepts of open data and public services.

The distinction between data(sets) and information seems to follow the general technical interpretation – for example, this is also mentioned in the Transparency and Privacy Review. However, in common language terms “data” and “information” tend to be used interchangeably. For clarity, it may be important in wider transparency communications (where a glossary is not available) to follow the common usage rather than the technical one.

On a linked note, the glossary seems to assume that the reader knows what the consultation means by “data”, which increases the potential for confusion between “data” and “information” since only the latter has an attempt at a true definition of the word itself. “Dataset” on the other hand, is primarily defined as “factual data” - which is somewhat circular.

If going as far to define what the consultation means by “data” – and by extension, what the wider transparency agenda considers to be data – then there seem to be two ways it could go.

- The first would to state openly that data is not defined by intention, and with it the practical scope of the transparency agenda. In other words, the definition of data would be entirely context-dependent and probably not far from “you know it when you see it”.
- The other possibility would be to attempt to define “data” rigidly, which could place restrictions on the scope of the transparency agenda - intentionally or otherwise.

Of these, the first seems preferable as it retains maximum flexibility. It also might help avoid the idea that “data” has to be numbers, when in practice there may be a range of helpful non-numerical data that could be released as part of the transparency agenda (for example, address details of brick-and-mortar front line government services).

#### *b) Where a decision is being taken about whether or not to make a dataset open, what tests should be applied?*

We believe that a fair test is encompassed in the framework described in our Transparency Implementation Plan. (Available online at [www.hmrc.gov.uk/transparency/implementation-plan.htm](http://www.hmrc.gov.uk/transparency/implementation-plan.htm))

Essentially, this tries to ensure that the benefits and opportunities of publishing data are considered fairly against the risks and costs of publishing data. This approach captures factors such as:

- The new default situation of publishing data unless there is a good reason not to do so. This should be the first consideration, i.e. actively applying the new default scenario in practice.
- The key risks from a breach of privacy or confidentiality (directly or indirectly) are then considered, as the only single factor that can override the default scenario of publication. The terms of the Commissioners for Revenue & Customs Act 2005 make it a statutory duty for HMRC to safeguard the confidentiality of its customers' data. This applies to both businesses and individuals.
- Other benefits should then be considered; for example the scope for commercial applications, the scope for aiding HMRC's role (the proper collection of tax and the proper collection of benefits, customer service, public accountability etc) or reflecting participation in wider government data releases.
- Other risks should then be considered; for example, whether the data could be used to enable fraud or non-compliance with the tax and benefit system, or whether data would have perverse behavioural incentives that would affect HMRC's operational capabilities.
- The resource impact of preparing data for publication should also be considered, although this should be about prioritisation (*when* it can be published) rather than *if* data is published or not.

Some datasets will have other considerations besides this basic list, so tests should be flexible to cover a range of situations.

Whatever the tests are, they should be consistent, credible, and transparent in themselves. Even the best such system would fail if it is done in the dark. Equally, it is important that where the decision is finely balanced, those demanding and supplying the data should both have a role in the decision. Otherwise, the decision will be made based on asymmetric data and one side may feel aggrieved unduly, which would serve to disengage them from the transparency agenda.

In other words the tests must be open and fair - and be seen to be so. This chimes with the notion of "metatransparency" discussed in the Dr O'Hara's independent Transparency and Privacy Review.

- c) *If the costs to publish or release data are not judged to be value for money, to what extent should the requestor be required to pay for public service data, and under what circumstances.*

As an extension of the reply to the previous question, the key underlying issue here is perhaps "judged by whom?"

As one of the objectives of transparency is to promote growth, then value for money must be judged by an objective cost-benefit analysis. The public sector is unlikely to have a full idea of the commercial benefits of a dataset, whereas the potential data users are unlikely to know how much effort and resources it will take the data provider to publish it. Without both sides bringing their knowledge together openly and a transparent way of arriving at the decision, the value for money judgement is likely to be skewed one way or the other.

Just as this process will help establish the value of the data, it will also help establish the resource cost.

We are aware that the Freedom of Information regime has the facility for a charging mechanism above a certain limit. Although this was developed well before the advent of the Transparency Agenda, the same rationales apply.

However, it may be incorrect to focus on the headline cost of providing data in money terms. What is likely to be of most interest to the data provider is the notion of opportunity cost – what would have to stop being done in order to publish the data for the first time and in the future?

If moving to publish the data has a direct impact on the ability of the data provider to also fulfil its core functions to the public, then there seems to be a stronger case for operating some form of charging system.

On the other hand, if the new data does not divert resources away from the direct provision of public-facing services or divert resources used to support those services, then there is much less of a case for any form of charging under the principles of Open Data and the Open Government License. The opportunity cost is also likely to vary from case to case even within a given organisation, so it is unlikely that particular categories of data could be brought into the Public Data Corporation's remit on grounds of opportunity cost.

If seeking to judge the true marginal cost of providing the data, then it could also be argued that the value of benefits to the public data provider should also be deducted (unless this takes the cost below zero). For example, the publication of some data may have a direct or indirect behavioural effect that supports the work of the public data provider. If so, then this should be accounted for in the costing – although this amount is likely to be very hard to evaluate in practice.

More thoughts on charging are at the start of section 2 in this response document: we believe that there are serious perverse incentives if the current charging mechanisms within the remit of FOI were altered as suggested in the consultation document, which could damage the wider Transparency Agenda.

- d) *How do we get the right balance in relation to the range of organisations (providers of public services) our policy proposals apply to? What threshold would be appropriate to determine the range of public services in scope, and what key criteria should inform this?*

Simplicity suggests that the range and coverage of Freedom of Information and transparency should be very strongly linked, particularly since transparency has been badged as "FOI 2" by the Minister for the Cabinet Office and in other circles.

However, we note that this simple definition would cause some tension with paragraph 5.30 in the Open Public Services White Paper, which suggests expanding the transparency agenda to also cover performance data held by those private companies and voluntary sector organisations providing (centrally-commissioned) public services.

e) *What would be appropriate mechanisms to encourage or ensure publication of data by public service providers?*

From the point of view of a data provider, the current dual approach of 'push and pull' seems to work fairly well.

The 'push' side of the mechanism could be said to be the centrally-driven system of departmental Business Plans and public commitments gives a clear focus for core public service providers and their various agencies and arms' length bodies, and has helped to drive the initial development and adoption of the Transparency Agenda.

The 'pull' side is perhaps the softer side of the mechanism, but has the potential to be much more powerful. This is more about facilitating the longer-term cultural shift to greater transparency as part of business-as-usual activities, and engaging with data users.

The work of the central Transparency Team, and the transparency boards (along with other players such as local digital engagement teams and transparency practitioners) should then support both sides as required. Having accountability for transparency rest at Board level within public data providers also helps ensure that the issue is given its due priority.

This arrangement also has the advantage of being quite flexible, and should be able to handle the vast majority of the range of issues required to drive and deliver the transparency agenda.

On a more tactical level, the messages given out by the centre also have a strong role to play – particularly in strengthening the 'pull' side. As with any cultural change, the involvement and support of the broader staff populations within public service providers is crucial to the effective adoption of transparency in the longer term.

This needs careful balancing; too little and the issue will appear to be unsupported and unimportant, whereas too much publicity only focusing on the positives will risk being seen as biased and could make it harder to convince people of the real benefits. If messages can help people answer the question "what is in it for me / my organisation / my customers" credibly, then encouraging action on transparency will be much more straightforward.

It is possible that (however well intentioned) using claims about the benefits of transparency that are hard to prove may do more harm than good in the long run. For example, the McKinsey report cited in the consultation document has some amazing figures for the potential benefits in respect of the public sector in the EU. However, the description of the methodology used to generate those figures is rather sketchy, and there is a lack of transparency in the finer details which means that readers cannot attempt to verify or duplicate the estimates. When those results appear to be based on some very crucial assumptions which may not hold true in the real world, and a potentially unrepresentative set of starting information, the power of the research is somewhat diminished.

This then leads on to the idea that public data must be trustworthy in order for the transparency agenda to succeed – please see our response about data quality (5e in this document) for more detailed thoughts on this topic.

## 2) An enhanced right to data

Before responding to the five specific questions posed for this part of the consultation, we note that there are several other questions listed in the main text of this section. (Part 4 of paragraph 8.6, which seems to focus on giving greater incentives to publish data under FOI requests.)

Taking these in turn:

- *Should there be a higher cost limit for datasets to that provided for other information under the FOIA so that more are released? If so, what should this be? Would the additional resource required be proportionate to the aim of increasing transparency?*

Firstly, this seems to link up to points raised about the consultation's glossary – given that FOI is enshrined in law, a much tighter definition of “data” is needed if aiming for separate cost limits for datasets.

That aside (and if we have interpreted this question correctly) setting a ‘reasonable’ higher cost limit for “data” under FOI legislation is unlikely to lead to vast quantities of extra data being published as a result of FOI requests or even lead to the right incentives in general.

It does not seem possible to set a well-informed higher threshold for “data”, as the range of possible data requests is so broad and the resource costs of producing new datasets can be very considerable given the steps that should ideally be taken (as set out in more depth in section 2d below).

It also seems slightly odd to focus on increasing the potential payment levels for new datasets when a key part of the Transparency Agenda is to encourage the free publication of data under the Open Government License.

Taking this a step further, higher charges for FOI data releases would also seem to strengthen other perverse incentives and undermine the general premise of the Transparency Agenda. This is because payment often implies ownership at some level. However, either:

(a) Ownership by the FOI requestor denies access to other data users unless they then request and pay for it separately. This is against the general idea of the open government license and the Transparency Agenda, leaving aside the Public Data Corporation;

Or

(b) After the data has been released and paid for by the original data requestor, in the spirit of the Transparency Agenda the public data provider should ideally publish it online so it can be access by everyone. A disclosure under FOI is a disclosure to ‘the world’ – which gives all other users a free ride on the original requestor’s payment. This in turn removes the incentive of the original requestor to ask for it on a paid basis in the first place, unless they are particularly altruistic.

Instead of setting higher cost levels for FOI requests involving ‘datasets’, it would seem to make more sense for there to be an open discussion between data users and data providers about the objective comparison of relative costs and benefits of publishing new datasets in general, as discussed at length elsewhere in this response.

We also note that there is already flexibility in the current limits in practice; we understand that public departments have discretion in the case of fees exceeding more than £600 (or involving more than 3.5 days' staff time). This can take the form of either charging the fees plus disbursements, declining to answer the request on grounds of expense, or simply providing the information free of charge regardless of the limit being exceeded.

- *Should a public service provider be required to pay some of the cost so as not to create an incentive to inflate cost estimates? Does the Information Commissioner provide an adequate avenue of address where costs are miscalculated?*

While this would undoubtedly be attractive to data users, a broader analysis suggests that this could be counter-productive to the Transparency Agenda overall.

Public data providers will already pay the cost of new data publications, whether through responding to FOI requests or preparing proactive data releases, or through the diversion of resources onto extra data provision activities. (I.e. there is an opportunity cost in terms of alternative resource uses, which will generally reflect on the delivery of other core support and/or front-line services.) Notwithstanding departmental commitments to the Transparency Agenda in general, there is already a strong incentive to optimise costs of data delivery in terms of a desire to also still be able to provide existing core services as well.

To make public data providers pay twice – expending resources and a share of the monetary cost on top – seems to do nothing except increase the total cost of providing data and decreasing funding for other areas of service provision (assuming fixed budgets).

- *If pro-active publication of datasets were mandatory, should a public service provider be able to refuse certain datasets on the grounds of cost – if so, what would be an unreasonable cost?*

Mandatory pro-active publication of data seems to be something of an oxymoron? Surely publication can either mandatory or proactive, but not both.

Interpreting this question as “should the FOI cost limit and associated right of refusal in the absence of payment be retained”: this again seems to go back to the notion of the opportunity cost of providing public data, discussed above. Presumably the current cost limit in FOI legislation is there to ensure that public bodies are able to safeguard a reasonable proportion of their resources for providing their core services in addition to providing requested information. This seems reasonable to keep, especially given the reduction in public sector funding as part of the austerity drive.

In terms of what an unreasonable cost would be: this is very likely to be dependent on the nature of individual datasets, and would have to be assessed on a case by case basis.

We suggest that the framework in our published Transparency Implementation Plan might be a useful model. (See section 1b above, in

particular, and the idea of being able to discuss relative costs and benefits openly between data users and data providers.) This may or may not involve migrating particular requests outside the remit of FOI – if so, such a move would have to be with the agreement of the requestor – if FOI does not provide for this level of discussion and flexibility.

- *Should a public service provider be required to publish datasets over the cost limit if the cost is met by the requestor? Are there circumstances where this may prevent the public service provider from delivering its core functions?*

Again, we have concerns over this idea. This is also largely based on the idea of opportunity cost but also on the 'stickiness' of the relationship between money and staffing levels.

On the first part: as discussed above, work on releasing data has a very real risk of being at the expense of other analytical support work for front-line services, if not at the expense of front-line services directly.

On the second part: being paid for a piece of work does not necessarily make it easier to provide it, due to the practicalities and economics of hiring or re-deploying staff.

Providing data often requires a degree of specialist statistical or other professional analysis skills, which means that there is not always a pool of the right people to re-deploy quickly. After a certain level of payment, it becomes possible / economical to fund new members of staff with the right skills to deal with the extra queries (ignoring any potential issues of recruitment freezes etc). However, below that certain level it is not optimal to use the payments for data to buy extra staff, and the payments may not do anything to aid the provision of data without cost to other public services.

A further practical matter is that requests will be on a range of different topics – using HMRC as an example, it could be about any of the taxes or benefits, or about performance, expenditure, staffing and so on. Analysts are generally specialists in each area, and it can take a while for new staff to get up to speed with the relevant topic or for re-deployed staff to learn the ropes of their new area. In order to justify paying for extra staff or re-deploying existing staff in practice, those 'certain levels of payment' would need to be concentrated in a particular subject area.

If paid FOI requests are spread across a wide range of topics instead, then it is much less likely that those 'certain levels' will be reached in any given area of the organisation. In this case, payment for data requests will not result in extra resource and capability, and so it becomes much more likely that data provision will be at the expense of other functions.

- *Should there be a new, higher cost limit for FOIA requests for data held within ICT systems procured after July 2012? Raising the cost limit would provide an incentive for public service providers to give due weight to the importance of designing systems from which data can be extracted quickly.*

We do not believe that this will have the desired effect.



Firstly, if IT systems of the future will be designed with the ability to produce data more efficiently (see section 2e below) then public data providers should be able to produce more data within the current FOI cost limits anyway.

Secondly, this idea appears to suffer from the same perverse incentives as the broader suggestion to raise FOI cost limits above – please see our response to the first bullet point above for more details.

Thirdly, this would seem to introduce greater complexity into the FOI system in general, and would require analysts and FOI teams to know exactly when relevant systems were procured before being able to judge which cost limit to apply. On a linked point, many datasets are compiled from relational databases drawing on multiple systems and sources, so there is a real risk that this rule would only apply to a fraction of an FOI request in practice.

Fourthly, in order to avoid transitional issues for new IT systems, the rule would have to be “procured *and activated fully*” after July 2012. If systems are procured but not in use by that time, then public data providers would be disadvantaged in the interim – a higher cost limit, but being stuck with old systems for a time.

Lastly, the ICT-contract-related suggestion elsewhere in the consultation paper would seem to be sufficient and much more efficient. (See our response to section 2e below for more detail.)

On the main consultation questions in this section:

a) *How would we establish a stronger presumption in favour of publication than that which currently exists?*

To clarify – “currently” for HMRC is that we have public commitments to upholding the [Public Data Principles](#) and the idea that we will publish data where possible. Based on the development of the practical side of the Transparency Agenda over the last year, we are now putting in place a full set of guidance material to facilitate this in HMRC for the longer term. Our published Transparency Implementation Plan also shows clearly how the default scenario of publishing data (unless there is a good reason not to) is embedded in our approach to transparency.

If these kinds of public commitments are insufficient then the next step would seem to be enshrining the presumption of publication and the Public Data Principles in law. If so, then this would create a level of legal enforcement and – presumably – a fair sanctions framework which would apply to those judged to be failing to uphold the law.

There is also an element of cultural change management to this. As detailed elsewhere in this response, it is vital that staff at all levels in public data providers are brought onboard.

This is likely to be much more powerful a lever than a legal solution in the long run. In general, change initiatives are much more likely to succeed if those affected are involved, allowed to help shape the solution, and brought around by clear and convincing arguments. Forcing change all the way through rarely achieves the intended benefits, and does little to engender a sense of support amongst

participants. (The latter can also mean that those involved are not motivated to design new innovative approaches, which the Transparency Agenda appears to require.)

- b) 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?*

Yes, this suggestion seems to be the most effective option.

Independence, objectivity and consistency are key features of an effective watchdog or regulator. Having a clear remit to compare, contrast and adjudicate fairly across different topics and different public data providers is also very important. The body's actions and investigations must also be transparent and open.

These characteristics are crucial in order to ensure that there is a fair balance between upholding the right to data and guarding against risks of sensitive data being released. The independent body could also help resolve situations where there is a tension between the right to data and the scope for particular datasets to be abused if released.

A sanctions framework to back up the independent body would also be useful in ensuring that its decisions were obeyed and enforced, but only provided the range of measures available to enforce its decisions are both fair and credible. (See section 4c below for more detailed thoughts on sanctions.)

Decisions by the regulator may not always be to the satisfaction of all involved parties, but it is vital that the independent body's decisions are impartial, trusted, accepted and implemented.

- c) Are existing safeguards to protect personal data and privacy measures adequate to regulate the Open Data agenda?*

HMRC is something of a special case in this regard. The terms of the [Commissioners for Revenue and Customs Act 2005](#) place a statutory duty on HMRC staff to protect the confidentiality of our customers' data. Contraventions of this statutory duty by disclosing information about an individual (whether directly or indirectly through a process of deduction) can be punishable by a jail sentence of up to two years and/or a fine. This penalty would apply to the individual member of staff responsible for the breach.

If anything, this legal duty combined with the very high priority placed on data security since the loss of two discs containing bulk data on child benefit claimants in 2007 means that the level of cultural change required to embed transparency is even greater in HMRC than elsewhere.

Staff are used to protecting data at all times, which provides a degree of a built-in safeguard. A key challenge for HMRC is to be able to refocus this level of protection so that it fits alongside the Transparency Agenda.

Our published Transparency Implementation Plan shows how we intend to achieve this, fitting the drive for publication of more of our data alongside the statutory duties to protect our customers' data.

*d) 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?*

The resource implications of an enhanced right to data should not be underestimated for the public sector as a whole.

In almost all cases, it is not true that data and/or information can be published at the press of a button. It usually requires a process of:

- understanding the new request or requirement
- collation (often from multiple systems and data tables)
- initial quality control and assurance
- scanning for potential disclosure issues, in line with our statutory duty to protect individuals' data (Commissioners for Revenue and Customs Act 2005)
- redaction or anonymisation as appropriate (or even going back to adjust the parameters of the dataset in its entirety)
- circulation to relevant stakeholders for further quality checks
- acceptance and sign-off (possibly involving a submission to senior management and ministers for more sensitive topics)
- drafting press lines for some releases
- setting up or adapting the relevant internet page
- linking to data.gov.uk
- dealing with any follow-up queries that publication might generate
- taking account of feedback in the design and content of future editions.

Taking shortcuts on this process compromises quality (see section 5e below) and greatly increases the risk that we would breach our statutory duty to protect individuals' information.

If the data in question could be classified as official statistics then there are additional steps to follow, in line with official statistics processes and the Code of Practice for Official Statistics.

At a time of declining public sector spending and resourcing, the additional burden of the transparency agenda has the potential to disrupt departments' core functions – if not directly affecting front-line services, then indirectly through the diversion of resources that would otherwise be used in support of frontline services, such as in analytical divisions and/or policy areas.

This applies to both requests for new data and enhancements of existing data. For example, we have measured the resource impacts of preparing the monthly

£25,000 expenditure data in HMRC for a recent informal consultation by HM Treasury. We estimate that at the current spend data threshold, it costs HMRC approximately 120 man-hours per month to publish this data; factoring in the mix of grades and standard hourly costs, this equates to around £4,250 per month.

One option suggested by HMT to improve the quality of the expenditure data is to reduce the threshold for all departments to £500, in line with that used by local government bodies. (Although strictly speaking this increases the quantity of data rather than the quality.) Analysis of our expenditure data shows that this would increase the volume of data in a typical month from 250-300 lines to between 20,000 and 25,000. Combined with other quality improvement measures already put in place recently such as extra descriptions of each expenditure item, we estimate that the resource cost per month of preparing this single dataset would increase to around 1,100 man-hours per month, at a financial cost of between £45,000 and £50,000 per month.

Extra resources are not available involved given the competing drive for austerity and the need to protect front-line service delivery, so these potential extra costs of transparency can only be met by stopping other core work undertaken by the relevant teams.

This single example is not to say that all new datasets we could release as part of transparency would be as expensive – although to balance this statement, many other datasets will inevitably prove to be more expensive, particularly if specialist statistical skills need to be applied to huge datasets in order to make them safe for release – but it highlights the key point that transparency is not free.

This is a point that does not seem to have been appreciated fully so far. Even if the costs can be absorbed without direct impacts on our other core activities, transparency work carries an opportunity cost in that the required resources would otherwise be deployed elsewhere. Any absolute increase in resource requirements will also not have been factored in to departmental settlements in Spending Review 2010.

Nevertheless, we are committed to the transparency agenda and will do all we can to drive it forwards. In practice this means that while the resource costs of transparency mean that we cannot release everything at once, resource issues will not be used as an excuse not to publish data at all. Instead, we will apply a prioritisation process based on the balance between resource cost, data availability and demand for data. This approach is set out publicly in our [Transparency Implementation Plan](#).

(See section 6b below for similar thoughts about resourcing and prioritisation.)

e) *How will we ensure that Open Data standards are embedded in new ICT contracts?*

This is probably more of a matter for IT and contract specialists, but a short analysis of the basic principles suggests something like this:

- Start by defining exactly how the standards map across to generic / typical ICT contracts.

- Then working with departments to pick up any special cases and practical implications; use these to refine the initial generic approach.
- Then develop a suitable range of standard or appropriately customisable text to use in public sector contracts.
- Failing that, create a set of clear and practical guidelines which set out the broad principles to apply.
- The wider principles may also be needed as a backup to any standardised text, to cater for the possibility for further special cases in future.

Taking this a step further, the enhanced needs of transparency in terms of data access, collation and reporting probably means that analysts and other (direct, departmental) data users also have a stronger role to play in the scoping and development of new IT systems. The owners of IT systems are unlikely to be the ones using them on a day-to-day basis to extract data for analysis and publication, whereas the analysts performing those roles are likely to have good input in terms of the practical uses of the systems' data and ways to optimise the systems' interfaces.

This probably means a much greater role for planning ahead in terms of the potential uses for the systems' data, which may be easier said than done in practice. There will usually be some level of forecast error, as all future uses cannot be known with full certainty ahead of time (just as we cannot predict with absolute clarity what uses data will be put to when it is released as part of the Transparency Agenda). There will also always be a need to strike a balance between the functionality of a system and its cost.

### 3) Setting open data standards

Before tackling the formal questions posed in the consultation document, paragraph 8.8 needs to be clarified. From the context, we believe that “quality” in this section refers specifically and only to the *format* and *openness* of the data (as per the five-star scoring system).

However, if quality is defined more broadly to pick up accuracy, truthfulness and trustworthiness of data as well, then we have serious concerns about this sentence in the consultation: *“Given the costs of improving quality, our judgement is that we should publish data of lower quality in preference to holding it back, while seeking over time to drive up the quality of that data”*.

We believe that placing too much emphasis on speed against the more general idea of quality poses a key risk to the successful adoption of the Transparency Agenda – please see our comments in section 8e of this response for more detail on this point.

Also, in terms of the suggestions in paragraph 8.9 of the consultation document:

- The first two bullet points are about commitments to the Public Data Principles, for which organisations can be held to account – these are already captured within commitments made in the HMRC Business Plan 2011-15. We understand that these were effectively mandatory elements for all departmental Business Plans across government.
- If seeking to make them more official through a code of practice (as mentioned in the first bullet point of paragraph 8.9) then it would make sense to create explicit links to the Code of Practice for Official Statistics to avoid duplication and confusion through having competing data-related codes of practice.
- We support the five-star scoring system as a very useful way of setting out the expected standards in terms of data format. However, we also note that we face technical challenges in terms of our ability to support linked data formats at present. For example, our experiences with setting up the interactive organograms did not go smoothly as colleagues in the Cabinet Office Transparency Team can attest – although one again we would like to thank them for their support! More generally, there are also skill and resource challenges with the more advanced data formats i.e. linked data – so we support the idea of releasing up to three-star data in the short term while aiming for four or five stars in the longer term.
- Just as an observation that links to the glossary section of the consultation, we note that the descriptions in Tim Berners-Lee’s five-star scoring system in places refer to data as “things” and “stuff”. If it means anything, this could be said to support a very flexible definition for “data”.
- If setting out how citizens can challenge where there is a failure in the process – this can only be done if “the process” is set out in detail. If it is based on misplaced expectations (on the part of either data users or data providers) without evidence that something is actually amiss according to the letter and the spirit of transparency, then there is unlikely to be a satisfactory resolution. Similarly, if there is an “open data compliance monitoring process” then there needs to be absolute clarity over what people are meant to comply with and how it will be measured. Both these

points link in with comments on an independent body to oversee transparency, as discussed elsewhere in this response.

- An obligation to at least consider feedback seems essential to the effective running of the transparency agenda. This is particularly the case if (as we have been informed several times during the course of transparency work so far) things such as data unlocking requests posted on data.gov.uk do not have the legal force of FOI requests and it is up to the receiving organisation to decide how or if they want to respond. This seems to be a recipe for very varied handling across government at the very least, some of which may not be in the spirit of transparency and openness. However, this then opens up further issues in terms of resourcing, so it is important to have a balanced framework for tackling data requests in general.
- As time and resource allows, we are happy to contribute in shaping the “general standards for data release” referred to in paragraph 8.10.

*a) What is the best way to achieve compliance on high and common standards to allow usability and interoperability?*

From the context, we assume that “compliance on high and common standards” refers to a general target for some point on the five-star scoring system.

If so, then the ingredients would seem to be the same as any good practice performance measurement system:

- clear targets
- a robust way of measuring how a given organisation matches up to those targets,
- support mechanisms to help organisations develop in line with any long term movement of those targets
- incentives for meeting those targets (either carrot or stick, or both)

This seems to have strong links to ideas about an independent body to oversee transparency, and about the potential need for a sanctions framework. (See sections 2b and 4c of this response document respectively.)

*b) Is there a role for government to establish consistent standards for collecting user experience data across public services?*

Some clarification would be helpful here. Does this question refer to the user experience in terms of open data, or user experience in terms of broader satisfaction with public sector services as a whole?

If it is the former, then this may be more a role for an independent body rather than the government. This would help ensure that transparency experience data is fully open and transparency in and of itself, and link to the “compliance on high and common standards” point addressed above.

If it is the latter, then the sheer range of public services suggests that it will be very hard to come up with meaningful consistent standards. There may be useful lessons from the recent development of the “Quarterly Data Summaries”, which

has attempted to gather a large range of data about the performance, spending and workforce of government departments within a consistent framework and using a (somewhat) standardised template.

c) *Should we consider a scheme for accreditation of information intermediaries, and if so how might that work?*

It would definitely be worth considering such a scheme, so as to analyse the issues in more depth.

“Information Intermediaries” do not seem to be defined in the consultation document, but the most obvious interpretation to us is “users of open government data who have created products and applications to add value to the raw data”. Applying this definition, the issues include at least the following:

- Accreditation usually signifies that an assessment of quality has been carried out against a set of standards, and the thing in question has met those standards. The first step would therefore be to construct that set of standards. (Ideally adapting something that already exists, such as the Code of Practice for Official Statistics?)
- Accreditation would also need to be transparent. Going by the definitions of datasets and information in the glossary, information implies some level of interpretation, potentially with some kind of agenda in mind. There is a risk that accreditation could then be seen as a way to further the causes of some agendas, while damaging the others. Impartiality would be key, which reflects back on the way the standards would have to be designed. The standards should also be consistent over time
- Along the same lines, a further risk is that accreditation could become a barrier to entry for new information intermediaries. If the first information intermediaries are the first to gain accreditation, then this would seem to give these ‘incumbents’ a degree of competitive advantage.
- More broadly – does an accreditation system for information intermediaries go against the grain of the idea of releasing data for free use in whatever way people see fit?
- There would seem to be a role here for existing bodies, such as the Government Statistical Service (GSS) and/or the Royal Statistical Society (RSS). These organisations already seek to promote the effective interpretation of data and statistics; for example, the Code of Practice for Official statistics includes a Principle which drives the production of effective and impartial commentary in support of statistical products, to aid users in avoiding misinterpretations etc. The Office for National Statistics also accredits official statistics products; those meeting the standards of the Code of Practice for Official Statistics are deemed to be National Statistics, and awarded a kitemark. The RSS has a programme to improve the level of education on the effective and proper use of statistics, particularly in the field of journalism.
- If the work of accredited information intermediaries is subsequently found to be wrong or misleading, how would that reflect back on the wider Transparency Agenda? This links in to the idea that data released as part of



transparency needs to be trustworthy; by extension, so should the uses of those data. (See section 5e below for more discussion of that point.)

- The previous point also seems to underline the need for safeguards in the system as well. If accreditation does confer some sort of advantage, then information intermediaries may not be entirely truthful in their attempts to become accredited. But if accreditation does not confer any advantage, would anyone go to the trouble of putting themselves forward for it?

#### 4) Corporate and personal responsibility

- a) *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.*

This question is probably not one that can be answered to best effect by a public data provider. We suspect that most other respondents answering this question will have different perceptions about whether commitments already given (in departmental Business Plans for 2011-15, for example) will be honoured. We also suspect that different respondents will have varying views on where the balance between openness, privacy and security should lie. It will be interesting to see the broader results for this consultation question.

Otherwise, we note that the final report of Dr O'Hara's Transparency and Privacy Review is very helpful in setting out a lot of the issues around this question. HMRC has a keen interest in the balance of these two forces, given our dual commitments to transparency alongside our legal duty to protect the confidentiality of our customers' data. A large number of points in that report mirror our own emerging approach to transparency, many aspects of which are set out elsewhere in this consultation response.

- b) *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?*

From the phrasing, we are not quite sure of the point of the first part of this question. In general though, personal responsibility (accountability?) at Board level seems necessary to give sufficient overall attention to transparency against competing priorities. However, there also needs to be lower level responsibilities for transparency at a lower level of the organisation in order to ensure that it is carried forward on a day-to-day basis.

It does not seem essential that those with overall responsibility/accountability for transparency and privacy are the same person. Having a single person act as the focal point for balancing competing views may be useful in some circumstances, but does not seem to be vital. That said, it is clear that there should be strong links between the two issues in governance arrangements if the two roles are held by different people, to ensure that there is an effective and balanced way of reconciling any tensions if transparency and privacy pull in separate directions.

On a linked note, it may be optimal for the responsibilities for sharing data and protecting data to be separated between different teams below the level of the senior responsible individual(s). This should allow the risks and benefits to be properly explored before bringing the two viewpoints together, helping to avoid the risk of an individual's natural bias one way or the other.

- c) *Would we need to have a sanctions framework to enforce a right to data?*

This seems to link strongly to the independent adjudicator concept discussed above. Sanctions could be helpful if targeted properly and if proportionate, but they

also need to be credible and fair as well. We would be interested to see what the balance on this question is between responding data users and data providers; in theory a fair consultation between the two sides should give something workable and broadly acceptable.

The key issue here is that of credibility if sanctions are applied. If data providers are perceived to be ineffective in applying open data, it is likely to be caused in part by one or both of resourcing issues (especially in tandem with austerity and spending review impacts) and privacy concerns. Sanctions are unlikely to improve either of these in practice, and financial sanctions would only make the resourcing constraint worse and could be counter-productive.

Reputational sanctions may be more useful in practice, as a way of giving an incentive for data providers to push transparency higher up the priority order when allocating scarce resources - although there will be some point where providers judge it to be sub-optimal to add further resource to transparency, given increasing risks to the effective delivery of core services.

However, a sanctions system should also have scope for an appeals process. Evidence coming to light in such a process could be helpful in ensuring that both sides of the argument can be seen. This could in turn act as an early warning sign, to help identify potential problems encountered in embedding transparency in the longer term before they become critical.

It is perhaps also worth considering sanctions on the other side of transparency as well. (For example, a small minority of Freedom of Information requests are deemed to be “vexatious”.) It can be argued that a fair sanctions framework should also have provision for coping with abuse of the right to data on the part of those demanding data as well as addressing failure to deliver by data providers.

*d) What other sectors would benefit from having a dedicated Sector Transparency Board?*

We believe this is more a matter for data users in terms of what they feel would be appropriate.

We do not have a Sector Transparency Board at present – we were not included in the initial round of Boards announced in July but we are happy to engage with the initiative if users of our data would find it helpful and proportionate.

## 5) Meaningful open data

### a) *How should public services make use of data inventories? What is the optimal way to develop and operate this?*

Our [Transparency Implementation Plan](#) was a fairly optimal way for us – balancing resourcing against detail, while still seeking to make it a workable core part of our approach to transparency. This suggests that there should be flexibility in the system for data providers to build something based on available information at the start, rather than being too prescriptive.

We also found the transparency-focused inventory work to be quite resource intensive, both for the central team constructing it and because input was needed from analysts and colleagues across the department.

Otherwise, we are aware of early scoping work by Cabinet Office colleagues assessing the feasibility central / standardised data inventory across all government based largely on descriptive fields similar to that in the data.gov.uk metadata. If this does develop further, then the similar lessons we learned about flexibility and resource cost should also apply. For example, many data providers may not have the required information for a central system already, so there will need to be scope for a phased and/or flexible approach?

For information, our Transparency Implementation Plan was published on 31 May 2011. Its web page received around 150 hits on the first day, and around 650 hits during June. After that, the hit count has tailed off markedly; fewer than 100 in July, and fewer than 50 in August. This is only one piece of evidence so extrapolations to the wider use of government data inventories must be treated with extreme caution, but it may suggest that - at the moment at least - data inventories are in the “worth doing, but may not be used very much in practice” category of transparency products. This might influence the level of resources devoted to them in future.

### b) *How should data be prioritised for inclusion in an inventory? How is value to be established?*

Our approach to our transparency-focused data inventory has been to:

- start with what is readily available (we had an old version of a data catalogue available in our analytical division)
- plus a trawl of the areas of HMRC were already plugged in to the transparency agenda as a result of those datasets mandated centrally (spending, performance, HR and so on).
- And be clear that the first version of the inventory section of our implementation plan is the initial version, to be updated over time as transparency is embedded more widely across the department.

This approach should be a relatively natural balance between the data which is readily available (even though a large proportion of it is subject to our statutory duty of confidentiality and cannot be released) and the data which is likely to be of most interest to users, in the medium term at least. Future efforts will then look to expand coverage and re-examine priority areas as required, especially in light of any feedback on our implementation plan.

We also found ourselves creating working rules for including or excluding data – for example, while HMRC-wide performance information is included, we did not go as far as listing lower level performance management data such as that collected by individual directorates. Similarly, we excluded many items of data that we use, but are owned or published (or not) by third parties. Since we do not own these datasets we cannot rightfully decide to put them in the public domain, and so we did not list them in our inventory.

However – and probably quite importantly – we have been very clear on the data types excluded from the data inventory, giving full details and explanations in our Transparency Implementation Plan itself. This is to ensure that our approach and decision-making is transparent as well.

In terms of establishing value – we believe that this has to involve both the potential users of public data and data providers in collaboration.

- Potential users will have a greater idea of the benefits and value of data in terms of commercial uses - albeit possibly subject to over-optimistic business cases and the difficulty in estimating the benefits in the first place. (As is perhaps illustrated by the McKinsey study cited in the consultation document.)
- Data providers will have a much better appreciation of the data quality and any privacy issues - albeit maybe overestimating the risk as a result of the cultural effects of previous data losses. That said, in our case at least the statutory duty to protect confidentiality is both strong and clear.

Without both ‘sides’ being aware of each others’ position on a given dataset, it will be very hard to reach a defensible and acceptable decision on the choice and/or priority of a dataset’s publication. In some controversial cases, there may be a role for the independent adjudicating body to step in, so that relations between data users and data providers (and their reputations) are not damaged unfairly in the longer term by particularly difficult disagreements.

*c) In what areas would you expect government to collect and publish data routinely?*

We do not see this as a particularly appropriate question for us to answer ourselves, but would be very interested to see the views of data users.

Our original informal transparency consultation on the transparency sections of November 2010 Business Plan 2011-15 suggested that (of responding stakeholders) performance and management information is quite high on the list. In our case, certain data from tax and benefit administration systems were also mentioned in our informal consultation, although there are significant privacy and confidentiality issues that would arise if seeking to publish much of the raw data. (This is the main reason why we focus on statistical products giving useful and non-confidential summary data.)

This also links in to the point above about the need for an effective link between users and providers of data.

Experience suggests that there is a reasonably big gap between the expectations of data collected by government and the reality of what is actually collected. A relatively common answer to FOI requests and Parliamentary Questions is that detailed data on many subjects is not collected (for example, as a result of the drive to reduce administrative burdens on businesses and individuals when dealing with the tax system) and so the question cannot be answered.

It is equally true that the public and other users of government data are not always aware of data that is already in the public domain. Another common theme of replies to FOI requests and Parliamentary Questions is to point to data that is already published, but where the requestors either did not know the data was already released, could not find it, or did not try to find it. To support this point, we note that so far the majority of comments received on HMRC issues on data.gov.uk have simply referred the commenting person to data already available. (The caveat is that we have only been sent a handful of comments or questions, so the sample size is small.) Data.gov.uk should go a fair way to helping resolve this over time, but it does highlight the importance of an effective knowledge management system for the UK - and the skills and willingness of people to actually use it - in addition to the publication of data itself.

*d) What data is collected “unnecessarily”? How should these datasets be identified? Should collection be stopped?*

Again we leave this mainly for data users to answer, and would be interested in the results of this question in the final analysis.

More generally, we are happy to support the work of the Government Data Review, and have a strong record of targeting and reducing overall administrative burdens for our customers. We will continue this drive to reduce the cost to our customers of dealing with us, and are developing performance measures for the costs of our individual customers to be published alongside our existing measure for the tax administrative burdens to business.

*e) Should the data that government releases always be of high quality? How do we define quality? To what extent should public service providers “polish” the data they publish, if at all?*

Where government data is concerned, “high quality” might be defined as data which is sufficiently accurate, complete and trustworthy (the three concepts are linked in practice).

Of these, the most important is probably the need to ensure that the data is trusted – without this, data meeting any other definition of ‘high quality’ will not be used and fail in its mission to inform. If anything, the drive for greater transparency may mean there is a need for even more attention to quality, so that a few high profile but low quality data products do not tarnish the reputation of the rest.

The definition of quality could also be expanded to pick up the context behind the data. A dataset may be perfect in and of itself, but if users do not have sufficient

background information to interpret it correctly then its *effective quality* will be diminished - or it could even become something actively misleading or dangerous. For example: table headings or methods of calculation may need to be explained in supporting material in order for users to properly understand the data itself. This turns 'quality' into something of a measure of 'usefulness' as well.

This approach mirrors the basis for the [Code of Practice for Official Statistics](#). For example, the foreword to the Code states that *"In a highly decentralised statistical service, the Code will serve to establish common standards and, by so doing, help to ensure a coherent and trustworthy service to the user of statistics."* This includes principles, practices and protocols on a wide range of data quality issues such as user engagement, impartiality, integrity, and release practices.

We are aware that work is ongoing in the Government Statistical Service and Cabinet Office to link transparency and the Code of Practice for Official Statistics, and think that this is very much moving in the right direction. Official Statistics status in itself - and the National Statistics tick-mark awarded to datasets assessed and found to comply with the Code in full - are very important signals of quality data in their own right.

In terms of "polishing" the data: we would be interested in what external responses to this consultation think "polishing" means - they may be indicative of the current levels of trust in government data as a whole. On the one hand it could be a relatively positive term, meaning that users think government is concerned with getting everything absolutely perfect; on the other hand users may see "polishing" as being more about massaging the figures to promote a certain point of view.

This is one of the key benefits of bestowing official statistics status on data; the framework puts a very strong emphasis on ensuring that 'government data' is impartial and free from political interference, and therefore more trustworthy.

We accept that there needs to be a balance between gold-plating a dataset and its timely release. And we also acknowledge fully that high quality datasets are more resource intensive, both in terms of the actual data preparation and from subsequent quality control measures, so there is a practical trade off between quality and quantity of data released given limited (and decreasing) resources.

However, we are concerned that a drive to release data without sufficient steps to ensure quality (and/or sufficient background information to explain the facts about data quality clearly) could easily backfire. This would damage the reputation of the transparency agenda *and* the reputation of government statistics in general.

Each dataset will be different, so a consistent framework will be very helpful for managing the trade-off between quality, quantity and speed of release. The Code of Practice for Official Statistics provides a good model for this – if not in full for all datasets, then a distillation of the key points. In particular, the principles of impartiality and user engagement should help ensure that data is both trustworthy and of sufficient quality to meet users' needs.

## 6) Government sets the example

### a) *How should government approach the release of existing data for policy and research purposes: should this be in a central portal or held on departmental portals?*

In general, we support the existing model of hosting data on individual departmental websites while linking through to data.gov.uk (either directly by people with suitable publisher rights, or indirectly via the feed from the Office for National Statistics' Publication Hub).

This has the advantages of being accessible through a single (government) search facility carrying standardised metadata, while also allowing public data providers to retain control of the actual publication process. The latter makes it much easier to maintain links, update data sources and so on. Our experience with data.gov.uk is that it is usually very easy and quick to link datasets, even if setting up a new page from scratch.

Centralising more of the process may just introduce further steps and/or stakeholders into the process of publishing data; if so, the complexity of the system and the risk of things going wrong would increase.

The current system seems workable and stable in practice.

### b) *What factors should inform prioritisation of datasets for publication at national, local or sector level?*

On the whole, the right ingredients seem to be present already – at least in theory if not in practice yet. Information on the demand for data and supply the supply of data needs to be matched before potentially published data can be prioritised.

Otherwise there are information asymmetries which can destabilise relationships between data users and data providers, and damage the effective functioning of the transparency agenda. For example:

- Data users are likely to have a much clearer view than the public sector of the potential benefits of publishing data - particularly the commercial benefits.
- Public data providers are likely to have a much better idea of the costs and risks of providing data, such as the confidentiality issues and resource costs (either in terms of direct expenditure or the opportunity cost from diverting resources from other work).

Linking this back to supply and demand:

- On the demand side: there needs to be the right ways to receive and encourage user feedback and requests for data. Data.gov.uk provides the facilities for this, as does the Freedom of Information legislation, Parliamentary Questions regime, and the new sector transparency boards.
- On the supply side: there needs to be sufficient information on the data available along with an understanding of the issues that publication might cause such as confidentiality concerns, and an idea of the resource costs of



providing new data. Data inventories should help meet the former point, while the involvement of data owners or data collators (e.g. statisticians) in the prioritisation process is required for the latter.

There then needs to be a way of pulling these two sides together objectively and transparently. For 'easier' data releases or those involving clear showstoppers (primarily involving privacy and disclosure issues) this process can be fairly natural and passive - much the same way as FOI requests are handled at the moment.

For larger issues and/or those with more areas of disagreement, then a way of actively bringing the two sides together may be required for effective resolution. This could be done through sector boards (provided these are sufficiently balanced) or through the formation of relevant bespoke user groups or consultation groups.

More generally (as also mentioned in section 2d above) it would be helpful if there was greater recognition that transparency is not a free activity for the public sector. It is not the case that there is a vast pool of untapped resource that can be suddenly devoted to publishing more data – especially as the drive for austerity means more focused resource usage in back-office functions, which are those areas owning, holding and analysing most datasets.

This is not to say that resourcing issues are an excuse to leave data unpublished, but it does put a limit on the amount of data that can be worked on for release at any one time. Transparency efforts have an opportunity cost, as noted elsewhere in this response. For this reason, we support strongly the idea of prioritising data for publication, so we can optimise the balance of our work to drive forward the transparency agenda as well as providing our other core services.

On a side note, analysis of Freedom of Information requests may reveal certain topics that would be suitable for pro-active release of data – this could also weigh into the prioritisation process. However, we recently undertook an examination of FOI requests about spending and procurement issues that we received over the last two years: this did not pick up any systematic trends in questions on that topic. The sheer range of procurement-related questions and the level of specific detail into which they tend to go into meant that no obvious candidates for 'FOI-forestalling data' could be identified for spending issues.

This experiment needs to be repeated on a wider scale to be sure, but at present the evidence suggests that pro-active publication of data may not actually lead to a significant reduction in FOI requests and parliamentary questions in practice (as is stated in the consultation).

Also, we hypothesise that some FOI requests and Parliamentary Questions are asked to highlight certain issues as part of a wider campaign, and/or asked so that requestors can demonstrate participation in the democratic process. Greater transparency will have a reduced effect on the future quantity of FOI requests and Parliamentary Questions if there are indeed other motivations besides the acquisition of data.

Lastly, there is the potential for greater transparency to generate additional FOI requests, Parliamentary Questions and correspondence. This risk could rise in to main ways:

- either people are unable or unwilling to find the data they want amongst the vastly increased amount of public data on the internet, and so resort to asking a likely-looking publishing body
- or the data published can be found and in the process increases demand for follow-up questions and additional data.

This is truly not to say that this is a bad thing – extra contact with users of data can be extremely beneficial, particularly in these relatively early days of the transparency agenda when seeking to prioritise our data for release and so that we can understand the external benefits of publishing data.

We simply raise these issues as part of a constructive challenge to the consultation's assertion that greater transparency will definitely reduce the cost and workload from FOI requests and other such correspondence, based on extrapolations of the evidence we have encountered over the last few years.

c) *Which is more important: for government to prioritise publishing a broader set of data, or existing data at a more detailed level?*

This may prove to be entirely dependent on the context of the datasets in question.

For example:

- More detailed data may be seen as being more useful.
- However, more detailed data is much more prone to raise issues of disclosure and confidentiality,
- A broader set of data can open up entirely new areas of information, such as publishing more types of performance data that can be of direct use to customers of the public sector.
- However, there is a risk that as the range of topics covered by data increases, it becomes harder it is to find any one thing - or even be aware of its existence in the first place. There is evidence from numerous FOI requests and Parliamentary Questions that even now, data users either cannot find or are unwilling to find data that has been published for many years.
- Publishing new data *may* be a little less resource intensive than preparing more detailed data - although there are arguments both ways. For example, new data is more likely be prepared by staff unfamiliar with the publication process and first releases tend to be more problematic in general, whereas those preparing an enhanced version of an existing data release may be able to tackle the process much more effectively as a result of greater familiarity.

Based on this analysis, we believe there should not be a specific central steer to publish either more data items over more detailed data or vice-versa.

Instead, it may be more effective to let this balance 'float' based on the combined views of data users and data providers, in line with whatever prioritisation framework exists (as described above). Then each case can be considered on its own merits, and should involve a more effective targeting of resources.

## 7) Innovation with Open Data

*a) 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?*

Given the (anecdotal reports of) slower than expected take-up of the use of open data, there does seem to be a role of some kind for government at this stage in stimulating its use. The most effective approach may be for government to create more of a supportive framework, but with sufficient restraint to avoid crowding out private sector users.

If government does too much to actually produce innovative applications itself, then there is the risk that (a) it takes up important space in the market and/or (b) spend significant resources on basic applications which soon become obsolete as other data users are inspired to produce something better. (The latter is less problematic in that it should be a sign of a successful kick-start, but it would make sense for resources spent on the upkeep of these sorts of government applications to be reviewed regularly and withdrawn once there are better external products available.)

Obviously, it also seems appropriate for government to focus on stimulating development of applications in areas that support citizens most directly, and in line with the central tenets of the transparency agenda (accountability, value for money and economic growth). Similarly - judging by the range of applications developed for mobile phones as an example, they vary from those which could be termed 'genuinely useful' to 'interesting' to 'fun'. Although there may not be a great many opportunities for the latter category when using government open data in the first place.

In terms of the practical approach, this role seems to link fairly naturally into the broader prioritisation framework discussed above. Even some joint brainstorming between data users and public data providers during prioritisation may prove to generate ideas for either direct government-developed applications or to kick-start other data users into producing workable ideas.