

## **DETAILED RESPONSE FROM HOME OFFICE TO QUESTIONS POSED IN 'MAKING OPEN DATA REAL'**

### ***Glossary***

#### **Q1. Do the definitions of the key terms go far enough or too far?**

The suggested definition of 'Dataset' is very broad and could apply to any random collection of data, or minutes of meetings and other documents. It is not wholly clear whether survey data is included. It is unclear whether the definition relates to both quantitative and non-quantitative data, and what 'factual' means i.e. whether attitudinal, opinion and subjective experience information are included or not.

The definition of 'dataset' probably goes slightly too far, and should explicitly exclude detailed information related to individual cases where information is provided for very specific purposes and its possible release [even if anonymised] would likely deter individuals from providing information. For example detailed information related to individual crimes which is provided in confidence (rather than counts of such crimes), or operational intelligence, should be excluded. Similarly we would suggest that in practice non-quantitative unstructured data, forecasts, progress against trajectories and scenario plans are not included.

The definition of 'open data' should clarify that the "presumption" is subject to the requirements of the Code of Practice for Official Statistics i.e. regarding both individual and statistical confidentiality disclosure issues, and also being subject to issues of resources, priorities and reasonable costs, balancing needs of users against burdens on suppliers and on respondents.

There are also additional terms which could usefully be defined, such as Open Licence, Open Source, Open Format, Transparency (and the difference between this and Open Data). It would also be useful to provide links to definitions of terms on other web sites where appropriate.

#### **Q2. Where a decision is being taken about whether to make a dataset open, what tests should be applied**

We suggest that the process of making a decision on the release of a dataset as 'open' should follow a process such as that outlined in the attached Home Office guidance on Opening Up Government Data. Important criteria include resources to release information evidence of demand from users for the data.

Proportionality should also be considered where the proactive release of data is being considered. The cost of provision needs to be measured against demand. Should we be spending money which could be used on frontline services to open data for which there is no evidence of demand?

We also need to consider if release of the data could be used inappropriately – for example there is concern that the release of the spend data could potentially be used to commit fraud. Therefore this should form part of the risk assessment when considering data for release.

#### **Q3. If the costs to publish are not judged to represent good value for money, to what extent should the requester be required to pay for public services data, and under what circumstances**

This option needs to be considered very carefully. Based on the experience of FoI we know that media corporations may see this as an alternative way of getting a story which we

would not have the opportunity to validate. We already have the public interest test in the FOI regime and this could be employed as a guideline for release.

We are already stretched under the requirements to deliver under FOIA. Whilst the presumption is that proactive release of data will reduce FOI requests we have not seen any evidence of this as yet. In contrast, FOI requests have increased year-on-year and we foresee that even under open data this will continue. What we do anticipate, based on experience, is that requests will become more complex.

We would strongly resist the option to pay for release of individual datasets as this would increase bureaucracy and inevitably impact on the department's ability to deliver its core functions. In addition, the release of data should be in the public interest not in the interest of the individual with the ability to pay.

However, where there is not a wider public interest in release, Departments should retain the ability to charge the full economic costs, or to make reasonable profits, in order to safeguard limited public resources. Equally there should be (as with responses to PQs and FOI requests) a minimum cost below which data should normally be released.

Departments should also be able to continue to refuse the release of data where

- i. requests are vexatious (e.g. multiple requests designed to obviate cost limits)
- ii. release is against the public interest (e.g. release of commercially confidential information which would then prejudice future government procurement, or release of early management information which would undermine the subsequent planned release of planned quality assured official statistics or release would prejudice frank and open recording and discussions of performance against trajectories towards government targets)
- iii. release would breach either individual confidentiality or statistical confidentiality rules
- iv. staff resources are not available or their use would result in significant delays to existing agreed, and high priority, outputs

**Q4. 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 criteria should inform this?**

The collection, processing and quality assurance of data, plus its publication potentially represents a burden both for the frontline and for sub contractors or data suppliers. Where data is already collated and supplied as part of the provision of a service then it should be in scope, if there is evidence of sufficient user demand. Data which is recognised as important in order to hold service providers to account will be published.

Where bodies are producing and publishing official statistics, these should normally be in scope for consideration to become Open Data. The National Statistician's report Identifying Official Statistics <http://www.statisticsauthority.gov.uk/national-statistician/ns-reports--reviews-and-guidance/national-statistician-s-guidance/identifying-official-statistics.pdf> provides a range of useful methods and criteria for deciding if data are official statistics or not.

Providers of services to local or central government (commercial or third sector) should not be expected to provide information on their own operations as open data which duplicates data they are already providing to local or central government as part of government statistical or research data collections, since government itself can then publish such information aggregated across providers as part of the accountability to taxpayers. An

exception is when such providers are mandated to publish the information as part of their contractual arrangements with government eg rail firms required to publish information by their regulator.

Any additional requirements for publication should focus on what public/developers are asking for.

**Q5. What would be appropriate mechanisms to encourage or ensure publication of data by public service providers**

In addition to building relationships with individual public service providers, including government departments, to encourage and ensure publication of data, it would also be useful to work more closely with cross-government groups, such as the Government Statistical Service.

Part of the development of mechanisms should include user contact and monitoring of use of data after it has been released – if it is not used or there is little or no evidence of value resulting from its use, then continued release can then potentially be stopped, thereby allowing resources to be focussed better.

***An enhanced right to data***

**Q1. How would we establish a stronger presumption in favour of publication than that which currently exists**

We believe that embedding “open by default” via existing legislation or introducing a mandatory requirement that all public bodies and providers of public services proactively publish data about the services they deliver would be disproportionate in terms of cost and resources in relation to public benefit. Building on the experience of releasing iconic datasets and the benefits that accrue from that is the way forward to moving to a presumption of release.

If proactive publication of datasets were to be made mandatory we would strongly favour the right, as a public service provider, to refuse publication not only on the grounds of cost, but also on the grounds of personal privacy and national security. The cost threshold should be the same as that as established by the FOIA.

There is also a need for centrally agreed and promoted guidance, rather than relying on each department to produce this locally.

**Q2. 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**

We understand that the review of the FOIA will cover this question more fully. However in terms of open data the current powers of the Information Commissioner are more than sufficient.

We are concerned that this approach is likely to increase amount of bureaucracy regarding the release of data. What has had most effect so far is clear leadership from ministers, especially the Prime Minister, and we need to build on this.

**Q3. Are existing safeguards to protect personal data and privacy measures adequate to regulate the Open Data agenda**

We believe that the existing safeguards to protect personal data and privacy measures are adequate, although more emphasis on this will be needed to ensure that the public

understand the implications of data being made available and open and that individuals consent to information being kept and possibly released about them.

Our other concern is around data being made available and what happens if it's not correct and is already in the public domain. What security measures are in place for the public if it is personal or sensitive data? Within the Home Office, we recommend that before the release of a dataset for the first time the checklist from the Privacy Impact Assessment is used to judge whether there are privacy implications in releasing the data.

**Q4. 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 its aim?**

The resource implications of this work are considerable, hence we need to focus on data that most benefits the public, and that the public wants. We need to approach Open Data in a sustainable way, otherwise it will become just another initiative that eventually bites the dust.

We also need to have a review date for datasets which are costing money to produce for which there seems to be no demand in the same way we have carried out the HM Treasury Data Review, and National Statistics review the collection and publication of official statistics. Reiteration from the centre that we should not be collecting data simply in order to publish it, but should focus on publishing what is already available, would be welcome.

**Q5. How will we ensure that Open Data standards are embedded in new ICT contracts**

We support the move to holding data in a way that it can easily be made open with little or no additional cost to the department under the Government ICT strategy. However we would need ensure that privacy and security are not compromised in doing so, and continuing guidance will be required from HM Treasury to ensure consistency of approach across government departments. We also would need to ensure that data released does not adversely affect our competitive advantage in securing value for money in contract negotiations.

**Other comments on this section**

On pages 24-25 you refer to incorporating Transparency requirements into procurement rules. In addition to this, the specification of IT requirements and system design need to be revised so that these elements are built in from the beginning. IT systems are the tools to deliver services: the business requirements and information management aspects, including Transparency, need to drive their specification and design, not the other way round. There also needs to be clarity in this section on the management of legacy IT systems.

***Setting open data standards***

**Q1. What is the best way to achieve compliance on high and common standards to allow usability and interoperability**

We do not see the benefits of having a formal Code of Practice which may become cumbersome and prescriptive in time: we consider that the broader scope of the Public Data Principles which are open to interpretation the preferable way to achieve compliance and develop common standards.

The five star rating for open data is helpful in providing a progression towards openness should be implemented where appropriate and where there is a demand or public benefit for providing data at 4\* and above. However, evidence from our statistical user community and

developer hack days has not provided any conclusive findings that there is a demand for data to be routinely released at the higher ratings, and the resource issues in undertaking this work suggests that this is best left to the developer community, although we should continue to actively engage with them regarding this.

We welcome the development to move away from separately held information registers over time. However it must be noted that the publication scheme is a legal requirement under the FOIA and any change to this will have to be mandated through amending legislation. We would welcome the opportunity to discuss the best way forward on developing a simple access list for the public – we do not want to see the development of a combined inventory that is cumbersome, costly to produce and ultimately out-of-date the moment it is published. It is also important that any work on standards-setting uses, or aligns with, other work on cross-government information management standards, and does not invent yet another set.

**Q2. Is there a role for government to establish consistent standards for collecting user experience across public sectors**

Yes. Through promulgating good practice / lessons learned from those lead delivery departments so that those departments which are less public facing can benefit from their experience.

**Q3. Should we consider a scheme for accreditation of information intermediaries, and if so how might that best work?**

Clarification is required as to what the role of these "information intermediaries" would be. Whatever this encompasses, we recommend that an existing professional accreditation scheme (eg Chartered Institute of Information Professionals, British Computer Society) is used if possible.

If none of these are suitable, then the cross-government Knowledge Council Professionalism Workstream could be commissioned to explore options, but if an 'in-house' scheme is adopted, then the benefits of this have to be balanced with the cost and resourcing of such a scheme. Other factors to consider will include who has responsibility for maintaining the scheme, and who will fund it. We need to avoid inventing yet another accreditation scheme if it's not needed.

**Other comments on this section**

The focus on releasing government data to date has been on new datasets. If the emphasis shifts to historic data then the cost implications of this are significant, especially if we have to digitise this to make it widely accessible, as at present much of this is only held in printed form.

***Corporate and personal responsibility***

**Q1. 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**

Within the Home Office we have found it useful to combine the role of the SIRO, who sits on the Home Office Board, with that of our overall Transparency Champion. A similar approach has been adopted in our Executive agencies and NDPBs,

We have also established a joint Public Sector Panel for Crime and Justice with the Ministry of Justice, which is chaired by Dr Kieron O'Hara from Southampton University, and which includes several, external and influential individuals from academic and business sectors.

This panel is proving an effective stimulus and sounding board for our work on crime mapping and justice outcomes, which are key to the Prime Minister's transparency commitments.

We also contribute to the main Public Sector Transparency Board chaired by Francis Maude, when their agenda renders this appropriate. However, we do not feel that an overarching Home Office Transparency Board, chaired at Ministerial level, could be justified in terms of the resources this would require from both Ministers and officials.

**Q2. 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**

The Home Office SIRO and Transparency Champion is a member of the Home Office Board, and is proactive in ensuring that both information assurance and transparency issues are considered with respect to data and information management. Within the Home Office we see these issues as being 'two sides of the same coin', and consistently stress the need for a balanced consideration of both issues. Our Statement of Internal Control for 2011, signed by the Permanent Secretary as Accounting Officer, included a section on Transparency in addition to the one on Information Assurance to further emphasise this joined up approach to the two subjects.

**Q3. Would we need to have a sanctions framework to enforce a right to data?**

There is no reason to suggest that the current framework is too fragmented, or that a 'sanctions framework' is required. We believe that it is much more productive to encourage and sell benefits of opening up data, and that a 'sanctions approach' is more likely to antagonise people, and make them look for loopholes. It is important to get people engaged, but also to provide strong leadership, as the response to the Prime Minister's letters in 2010 and 2011 has demonstrated.

Government departments should be able to work closely together – recent experience has shown that the Cabinet Office has perhaps not consulted as early and widely as it should have with others with responsibilities in the same field. We hope that this will change in the future as we all work together to make open data a reality.

**Q4. What other sectors would benefit from having a dedicated Sector Transparency Board?**

A review of the dedicated sector transparency boards should be carried out after their first year to see what benefits they have provided and whether the model is sustainable in the long term or of use purely as "pump priming" measure before considering rolling this approach out more widely.

***Meaningful open data***

**Q1. How should public services make use of data inventories? What is the optimal way to develop and operate this?**

We welcome the development to move away from separately held information registers over time. However it must be noted that the publication scheme is a legal requirement under the FOIA and any change to this will have to be mandated through amending legislation. We would welcome the opportunity to discuss the best way forward on developing a simple access list for the public – we do not want to see the development of a combined inventory that is cumbersome, costly to produce and ultimately out-of-date the moment it is published. Although the Home Office has published a redacted version of our Information Asset

Register, the full version includes some datasets which cannot be declared. It is difficult to envisage the situation where one list can meet all requirements.

We would welcome evidence that shows that data.gov.uk is the “portal of choice”. Our evidence within the Home Office does not support this – with the users of our key datasets either coming to us via Google or the UK Statistics Authority Publication Hub. We believe that data.gov is not achieving the hit rate it should if it is to become the way the public is to access government datasets. We would be interested to hear of the evidence from other departments on the channels that the public use to find their datasets.

The quality of data on data.gov.uk needs to improve – we have raised the issue of duplication of datasets where they are taken from Data4NR and the UK Statistics Authority Publication Hub on a number of occasions. Also the issue of defunct sources such as DCLG Floor Targets Interactive: for Home Office datasets this has led to a confusing situation, with out of date datasets able to be retrieved by the public, which we are unable to remove. There does not appear to be the resources to maintain a public inventory to the standard that would provide the best solution for the public.

**Q2. How should data be prioritised for inclusion in an inventory? How is value to be established?**

We support the proposal to develop inventories in a modular way over time. Priority data should be determined by the provider – they will know best what data is the most helpful and significant for their area. We see value in some central guidance on how to assess high priority data.

**Q3. In what areas would you expect government to collect and publish data routinely**

HM Treasury have recently carried out a review on this issue, which we understand is being taken forward by ERG in the Cabinet Office.

**Q4. What data is collected unnecessarily? How should these datasets be identified? Should collection be stopped?**

This issue was also addressed by the HM Treasury review, and is being taken forward by ERG in the Cabinet Office.

**Q5. Should the data that the government releases be always of high quality? How do we define quality? To what extent should public service providers 'polish' the data they publish, if at all**

We should build on the work of the HMT Data Review to establish quality. The Home Office's work on improving data quality was recommended. The release of poor quality data damages the reputation of the government and ultimately undermine the public's confidence and trust in government data. Good quality data does not mean “polished” data. Data releases should always be robust and fit for purpose – but politicians, civil servants and the public need to understand the risks of mistakes/misunderstanding of data quality, and the appetite to manage this risk.

We also need to be clear on how the public will be able to make a judgement as to the accuracy, currency and validity of the data they are accessing, particularly where data has been re-used and re-worked by organisations other than the authoring body.

We welcome the emphasis, on pages 19 to 21 of your consultation document, on the use of benchmarking to drive up the quality of data, and believe that this is particularly important in

order to ensure its consistency, cogency and that it is understood in context. We would welcome the opportunity to participate further in this work.

### ***Government sets the example***

#### **Q1 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?**

There needs to be a balance between usability for public, and certainty that they have found all data available, with ownership of data on part of departments, plus the ability to see the context in which data has been collected or created. The current position with data.gov.uk as the principal portal for the public, but fed by links from departmental websites provides this balance. Members of the public should not have to know what department is responsible for a certain subject or dataset to be able to find it. However, the searchability of data.gov.uk and the quality of data on this portal needs to be improved, in particular with regard to the duplication of datasets (see response to Question 1 in the Section on 'Meaningful open data', above).

#### **Q2 What factors should inform prioritisation of datasets for publication, at national, local or sector level?**

This depends on the purpose for which the datasets are being released. Is it released for the benefit of the public in knowing what their council or government is doing and what services are available or is it to facilitate the re-use and repurposing of the data by enthusiasts or the commercial sector? Logically the first should be more important at a local/sector level and drive the prioritisation there, with a more strategic approach being taken at a national level.

Ideally public demand for data, or requests by developers, but then may take time to respond if data not being collected in suitable way, so this approach needs to be combined with active consideration by departments of the data held and how could better make this available.

#### **Q3 Which is more important: for government to prioritise publishing a broader set of data, or existing data at a more detailed level?**

In terms of the underlying data behind policy advice and decisions the priority should be to build on publishing a broader set of data – decisions are taken in a broad context.

#### **Other comments on this section**

We broadly agree with the proposal to routinely publish evidence and datasets behind Policy Statements. This is already done by the Welsh Assembly and is a logical progression from the treatment of statistical information used to provide an informed background to the taking of the decision as set out in s35(2) of the Freedom of Information Act 2000. Arguably, we should go further and say that Government Departments should be required to create Policy Statements in a way to facilitate their publication. There may however, be certain policy statements which cannot be published due to security considerations.

### ***Innovation with open data***

**Q1. 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?**

Although government should not be determining what is done with data, we could stimulate innovation, for example in making grants to developers, highlighting applications that it would find useful (or perceives would be useful to public), and building collaborative relationships with both developers and informed individuals in the academic and commercial sectors. This activity could usefully be reported on in the transparency section in departmental annual reports in order to provide transparency on this work and to share best practice and experience.