

Making Open Data Real: A Public Consultation ACPO Response

INTRODUCTION AND GENERAL OBSERVATIONS

- ACPO welcomes and embraces this Government initiative which provides an interesting opportunity. There is a very real possibility that it will bring a different view and context to the interpretation of police information without the cultural history of traditional use. However, data requires context to make it valuable to users, and the maturity of the receiving audience is an issue.
- Information does not speak for itself; it needs to be carefully interpreted. A concern about the open data initiative is that it could lead to an increase in ill-informed and possibly vexatious questions and comments. In answering these, public services risk appearing defensive. Experience of intra- and inter-service comparisons has shown that it is dangerous to presume that variables share definitions and that consistent counting rules have been applied.
- As part of this agenda, and of specific significance for the police, 27 October 2011 will see Her Majesty's Inspectorate of Constabulary launch the new 'Crime and Policing Comparator' through the police.uk website and HMIC's website. This will supplement the existing crime mapper application that is hosted on the police.uk website by providing the public with various data sets (crime, cost, workforce and service quality) which they can then use to compare police forces' performance.

HMIC will not interpret or contextualise the data released via the Crime and Policing Comparator. Instead, forces will be at liberty to prepare separate statements to contextualise the data that will be linked to the HMIC website. Only by contextualising open data will it become truly valuable to users, and minimise the threat of reputational harm to organisations and damage to the relationship with the public.

- No real key definition exists in terms of the difference between data that is part of the transparency agenda and data sets which are discussed within this consultation paper.
- The paper also raised the question of whether the police service already has a process through supply of information to the Home Office Data Hub (albeit that it is still the subject of refinement) that produces suitable public information in a format that strikes the right balance between informative information and privacy of the individual. The police service already has over 110,000 publication schemes, and one would have to question whether this is already stretching the capability of the Information Commissioner's Office (ICO) and their governance role. The issue around applying tests with regard to data sets is a critical one which goes beyond privacy / publication issues and impacts upon data already provided to Government. There should be clear public benefit with regard to the depth of the data set made available.
- There is no articulation of the expectations of the public. There is a high risk that what will be made public is not necessarily what the public want but what public authorities have the capability of producing. A good example of determining what it is that the public want would be an analysis of crime mapping. It is prime example of making information available to the public. However, in data set terms, a company who installs burglar alarms might argue that the data they require should be home-specific as they would then be able to offer services which might protect that individual more effectively. In this case it is easy to define the difference between public and private. However the complexity becomes more exposed if crimes are linked to outcomes. This initiative is currently being reviewed by the National Police Improvement Agency (NPIA) with the intention that the result

NOT PROTECTIVELY MARKED

provided would only indicate against a crime that an offender had been convicted and the sentence that was imposed. This is unlikely to meet the aspirations of those who would want to show a photograph of the offender and their name. It is this kind of issue which sits at the very heart of making data available and ensuring that the balance is reached between privacy and accountability to the public.

- The solution to getting this balance right should be focused on public service solutions which are delivered in the local environment. For example, if one takes the MPS then within the boroughs of London there should be a framework approved through the Mayor's office which provides data that is relevant to the local public. This may mean linking bus timetables to hospital opening times to crime and disorder data to local education. The concept behind this is that a new resident in this borough wants to make a decision as to how her disabled child will travel to school, obtain access to local health treatment and live in an environment where they will not get bullied. This linked data approach provides a rich horizon of information that really adds value.
- As the Regulator, the Information Commissioner cannot be too closely involved in defining what data should be made public as they may well have to adjudicate on any complaint that is made. One may have to question whether a regulatory body who have to adjudicate on complaints can also be the architect for leverage of more data into the public environment.
- Some of the benefits do not necessarily apply to the Police service, e.g. you cannot choose your policing provider!
- Delivering Open data is going to be challenging particularly in the current economic environment.
- The scope of the consultation is extremely broad and non-definitive and therefore difficult to comment on constructively. We appreciate the desire to encourage creative thinking but it might be more productive to identify key areas where value would be added and accountability increased, and focus on getting them right first.
- Most of the police service's systems are relational (i.e. sets of linked data tables). This means that data sets could only be generated by running database queries and would be snapshots of a particular point in time.

QUESTIONS FOR CONSULTATION:

1. Do the definitions of the key terms go far enough or too far?

The definitions of the key terms in themselves are clear. However, as previously mentioned, there is an issue over the key definitions as to the difference between information under transparency and data sets.

2. Where a decision is being taken about whether to make a dataset open, what tests should be applied?

There will need to be incentives for organisations to provide open data (stick vs carrot) whether this is done by the ICO or elsewhere. Existing safeguards ought to be sufficient to protect personal data but no-one yet knows what the impact will be when data is aggregated. It is likely to be costly. The Transparency requirements and the publication scheme are already placing financial and resource burdens on police forces, and it is doubtful that pro-actively publishing the information will reduce the number of Freedom of Information Act (FOIA) requests.

NOT PROTECTIVELY MARKED

Standards for interoperability are going to be a huge issue. Trying to get common agreement across the whole public sector will be even more difficult than getting agreement across Whitehall. A narrower focus to start with would be helpful.

One approach would be to publicise data and set low expectations with the public in terms of the level of interoperability of data, and then use that to drive convergence i.e. put the data out as it stands and then someone is responsible for making it converge.

3. If the costs to publish or release data are not judged to represent value for money, to what extent should the requestor be required to pay for public services data, and under what circumstances?

Despite eight years of FOIA, it does not appear that openness and transparency are top of organisations' priorities. If the Government is serious about this, it will have to think seriously about culture change. The same person in each organisation should be responsible for disclosure and safeguarding of information so that a balanced decision can be made. A proportionality test should be applied to assess whether the cost of providing the information is "bearable" to the organisation. Organisations should be given the right to reasonably refuse a resource intensive request on the grounds of affordability. Fair "pricing" of information should be set against a matrix of charges that is agreed in advance and is acceptable to the organisation. A published scale of charges may reduce the number of requests. Open data requests should not be seen as an income generation scheme.

An alternative option to be considered, as noted later in the document, is that the Government publishes every data set it legally can do, taking into account other legislation, in open format and with limited attempts at formatting etc. The quid pro quo is then that a realistic charge is made for FOIA responses which are in effect an interpretation of published data.

4. 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?

There should not be an expectation that every public service will publish every set of data it owns. We are unsure how you could decide what would be the most useful data – there would be so many different agendas – but suggest it should be those with the widest potential for re-use. Data quality will be an issue – there will be lots of league tables generated that are not necessarily comparing like with like. Some organisations may be reluctant to release 'unpolished' data.

People are already drowning in information so a central portal to signpost is essential. Possible factors for prioritising: least cost, most value, most common FOIA requests.

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

We do not think there is a need to stimulate innovation – journalists, academics and businesses will be itching to get their hands on the data and do creative things with it. There will be a learning process whereby regular requests for data sets could be then decided as being a pro-actively published set to minimise the resource burden of separate requests.

SIX POLICY CHALLENGE AREAS - QUESTIONS FOR CONSULTATION:

The following set out our views in response to the wider questions asked in the consultation.

1. An enhanced right to data: how do we establish stronger rights for individuals, businesses and other actors to obtain, use and re-use data from public service providers?

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

We believe this would be best achieved through the creation of an additional Code of Practice, to sit alongside the existing Lord Chancellor's Code of Practice on the Management of Records (under s46 of the FOIA).

ii. 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?

Enhancing and extending the existing powers of the ICO would be most effective. The ICO already regulates the FOIA, Environmental Information Regulations (EIR) and Data Protection Act (DPA) and has an excellent track record of balancing the tensions between these legislations, and providing a well-embedded service to both the public and private and public service organisations. It would be more sensible to increase the powers of the ICO where skills, experience and knowledge already exist rather than create yet another body (as has already been done with the CCTV Commissioner) at additional cost to the public purse and run the risk of conflict with the ICO, which could be detrimental to the needs of the public.

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

Since the commencement of the FOIA in 2005 many public services (in particular the Police Service) have been making data available via their Publication Schemes and servicing FOI requests whilst preserving privacy and protecting personal data, as well as conducting reviews in compliance with the FOI Act. Therefore, if adhered to properly and if not confused or contradicted by new bodies, new principles, etc., the existing safeguards should be adequate.

iv. 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?

Firstly, cost limit. The current cost limit within FOIA is adequate and protects public services from frivolous requests and ensures that information disclosed is adequate and sufficient. Raising the cost limit will mean that more requests would need to be responded than currently, which would require additional resources within the public services, adding significant costs and increasing the current burden.

Secondly, time limits for reviews. The current ICO recommended timescales are more appropriate than legislative ones. The danger with statutory timescales for reviews is the potential for loss of quality, which may not be in the public interest. Alternatively, there is the risk that the reviews would take precedent over the normal business of the organisation, reducing service delivery elsewhere.

v. How will we ensure that Open Data standards are embedded in new ICT contracts?

Procurement can be a lengthy process and therefore the suggested deadline of July 2012 may be a little premature. It needs to be more reflective of procurement processes to provide sufficient time for the public service to implement this into the procurement process. Requiring that systems also have an efficient retrieval mechanism for data-sets for public consumption, could increase the costs public services would pay for their ICT systems and the time take to procure them.

Whilst it is unfair and unjust that requests for information can be refused due to ancient records

NOT PROTECTIVELY MARKED

management methods, the standards for ICT procurement would not address all the issues. Effective ICT provision is an end-to-end process and cannot be solved purely by the procurement process. Neither would this take into consideration legacy ICT systems or manual filing systems, etc.

Currently, public services are not required to create data purely for the purpose of publishing. Any change to statutory requirements should continue in this spirit.

Moreover, it is unsafe to presume that variables in different systems that share the same name also share the same definition and counting rules.

2. Setting transparency standards: what would standards that support an enhanced right to data among public service providers look like?

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

The proposed Public Data Principles could be confused with the Data Protection Principles. Some of the principles appear to be promoting the need to publish information speedily despite information being inaccurate. This may be acceptable in limited circumstances and provided the necessary safeguards are in place, however this appears to be in conflict with section 4 – Meaningful Data and point (v) regarding quality.

ii. Is there a role for government to establish consistent standards for collecting user experience across public services?

Data does not speak for itself. It is our view that metadata, interpretation context must be a key requirement for all data made available through the transparency agenda. Without clear and proper context, we would be likely to see users constructing inaccurate interpretations and inferring erroneous meaning from the data. This could lead to confusion amongst data users, and in the process undermine public confidence and the relationship between organisations and service users.

In addition, we foresee a risk that when using open data to compare and benchmark between service providers (e.g. in areas such as the effectiveness of service delivery and value for money), the absence of context would lead to inappropriate comparisons being made between them. This in turn could lead to undue public concern and unnecessary scrutiny of services. This would significantly undermine one of the key ambitions of the transparency agenda, namely to promote 'choice' in service provision.

Therefore we support the notion of setting data standards and definitions and welcome the Governments desire to work with data providers to set standards around data. The consultation paper states: *"For specific datasets, like spend or staffing data, we will also set data definitions, to explain exactly what the data mean, so that they can be combined and compared for re-use"* (p.27). We would encourage that definitions and context be applied to data as widely as possible, and that these should be determined in close consultation with services.

But, regarding the suggestion of *"having in place an Open Data compliance monitoring process which outlines how, when and where public service providers should report their progress"* (p.27), there could be potential for this to become a new bureaucratic process, which would ultimately work against the spirit of the Government's aim to reduce public service bureaucracy.

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

NOT PROTECTIVELY MARKED

Whilst we sympathise with the notion of considering and acting on user feedback, we have some caution around the suggestion that this should be the case “*even where it has been collected independently of the public body or public service provider*” (p.27). If this were to proceed, the Service would need to be convinced that strong checks and balances were in place to ensure that the independent bodies’ data and analyses was sufficiently robust and free from error and agenda bias.

3. Corporate and personal responsibility: how would public service providers be held to account for delivering Open Data through a clear governance and leadership framework at political, organisational and individual level?

i. 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.

ACPO already play a major part in relation to information rights within the Police Service and assisting with good practice standards. The ACPO Central Referral Unit collates requests from all forces and offer advice. The ACPO Portfolio for Data Protection, FOIA and Records Management discuss matters arising from the legislation and promote good practice.

Within each Force, support and backing by Chief Officers is essential in order to promote Open Data and ensure that information can be readily accessed and retrieved without hindrance and with adequate resources to undertake these tasks. The Government may consider introducing a training programme or occupational standards for leaders.

ii. 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?

Yes, this needs to be joined up as there is a direct relationship. Similar to the ICO, the openness needs to be balanced against protection of data and individual rights.

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

We already do, within the FOIA and DPA.

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

4. Meaningful Open Data: how should we ensure collection and publication of the most useful data, through an approach that enables public service providers to understand the value of the data they hold and helps the public at large know what data is collected?

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

Data Inventories are already in place in the form of the Information Classes within a publication scheme. For instance, the ICO sets out what information should be made available within each Public Service’s Publication Scheme to ensure that the information is adequate and necessary for the public. This ensures that unnecessary resources are not used to publish information that is not relevant or which does not enhance the public’s knowledge of the organisation. Therefore, similar agreement will need to be made as to what these inventories will cover.

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

NOT PROTECTIVELY MARKED

Research of current trends and issues will be required to ensure that the information is meaningful to the public to assist with the reduction of FOIA requests. It is impossible to publish every data-set (in respect of the Police Service) and to pre-empt the needs of the public.

- iii. In what areas would you expect government to collect and publish data routinely?***
- iv. What data is collected “unnecessarily”? How should these datasets be identified? Should collection be stopped?***
- v. 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?***

In the current climate there is an emphasis on only publishing data we are confident is consistent, accurate and complete. There are many examples of public services being criticised where this data is subsequently shown to be inaccurate or are not comparable with other data-sets. Whilst in principle we support the drive to make more data available, we feel there would need to be some safeguards in place to reduce the likelihood of such challenges if we are encouraged to publish data even where we believe it to be less accurate. There may be cases where publishing less accurate data would not be in the public interest. The Government would also need to ensure that public services do not deliberately supply erroneous data and wilfully mislead the public.

It is likely to be the case that data held and released by providers, even within the same sector/ service area, will vary in quality. This would, unknowingly, affect users’ interpretations of the data, and consequently lead to inaccurate conclusions being derived from analysis. In turn, this would mean that inappropriate decisions and/ or action could be taken as a result.

The paper also proposes that *“Given the cost of improving quality, our judgement is that we should publish data of lower quality in preference of holding it back, whilst seeking over time to drive up the quality of data”* (p. 26). For the reasons expressed above, we would not support this approach. Instead we feel that data (which, under the transparency agenda is intended to promote democracy, accountability and service improvement) should only be released when there is satisfaction with its quality. Releasing poor quality data could hamper confidence in services and relationships with users. In terms of defining quality, this needs to take account of accuracy, timeliness and context. As stated earlier, we feel that data without context is essentially meaningless.

We would support an approach where all data released should be carefully defined, with standardised definitions, and linked to other relevant data wherever possible. This would enable users to manipulate and analyse the data with a clear understanding of its meaning, allowing them to reach more appropriate conclusions through their analyses. Furthermore, this would facilitate much more accurate benchmarking between peer organisations, enabling this to be approached with greater confidence.

There is also something here about having an understanding of what data accuracy actually means. Granularity could be interpreted as a measure of accuracy but it is not. Care must be taken to talk about accuracy in the context of granularity.

We also need to understand provenance. How was data created and for what purpose? If something is measured for a specific purpose it will be measured in a particular way depending on the need.

- 5. Government sets the example: in what ways could we make the internal workings of government and the public sector as open as possible?**

We support the government’s objective to reduce the volume and demand of Freedom of Information requests on public services. The paper states *“If more data is made proactively available,*

NOT PROTECTIVELY MARKED

administrative burdens involved in answering requests may be reduced” (p.33). However, whilst the volume of Freedom of Information Requests for data may decline, it could be the case that the release of data actually generates more interest in organisational activity and consequently an increase in requests for further information. For example, where there is no immediate information to describe data context, this would likely lead users to seek more information from a service provider to clarify meaning. This could lead to a scenario where services are increasingly required to answer or indeed pre-empt questions, potentially increasing bureaucracy.

- i. ***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?***

In respect of the Police Service, data is currently published in publication schemes on the internet.

- ii. ***What factors should inform prioritisation of datasets for publication, at national, local or sector level?***

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

6. **Innovation with Open Data: to what extent is there a role for government to stimulate enterprise and market making in the use of Open Data?**

- i. ***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?***

In general we support stimulation of business, especially for local providers, and feel that the Publication Scheme is the appropriate route for this. However, we are concerned about the resource implications that encouraging additional FOI requests from the commercial sector may require.

We note the suggestion whereby public service providers could be required *“to report each year on how they are building collaborative relationships with the user community, including the commercial sector, which promote use of data”* (p. 35). We feel that caution should be taken around the introduction of reporting frameworks, as this could lead to structures and processes which generate bureaucracy.

We are sympathetic to the notion of external collaborators using our data to promote problem-solving. For example Greater Manchester Police are actively pursuing collaborative, problem-solving opportunities with research specialists by coordinating access to our people and data through an academic research hub.

We feel that innovation would be best encouraged and promoted through incentives for service providers and users, alongside the promulgation of good practice and examples (case studies) of successful outcomes. These would need to clearly demonstrate the business benefits that were realised.