

Cabinet Office - *'Making Open Data Real: A Public Consultation'*
South West Observatory Response

A. Overview & Background

A.1 The South West Observatory (SWO), as an established network of researchers, analysts and data practitioners working at local, regional and national geographies, values the opportunity to contribute to the public consultation on *'Making Open Data Real'*.

A.2 The SWO Core Unit has sought the views of the SWO network in compiling this consolidated network response. This response benefits from specific input from SWO's Housing and Planning Module (provided by the Homes and Communities Agency) and local network of intelligence partners (working mainly within research teams in Local Authorities).

A.3 For a number of years SWO has taken an active role in improving the provision and access to public data and information, including advising businesses on developing interoperable data hubs.

B. Summary

B.1 This response is supportive in favour of the principles behind Open Data and opening up public access to data.

B.2 It seeks to answer the questions posed by the consultation relating to scope and mechanics of encouraging open data, using local examples/case studies where possible.

B.3 However, it raises two main issues in need of further consideration, namely data quality and the needs of the user. It also challenges some of the assertions implied within the consultation, specifically that the process of opening up public data can reduce the administrative burden on services; that the benefits to publishing datasets, even if quality is not high, always outweighs restricting its release; and that, as a result of more Open Data, FOI requests should tail off, thus saving money.

Questions for Consultation

C. Glossary of Key Terms

Questions from p5 -

C.1 There were no major issues raised in relation to the key terms used within the consultation. However, it was noted from the outset that Open Data does not necessarily imply 'Open Information'. For information to be reliable and valuable it needs to be analysed in context and with expertise. This represents a potential risk to the whole Open Data agenda and the value that it brings.

C.2 In terms of tests that should be applied in making a dataset open or not, replies suggested that unless data would cause a danger to people or commercial investments it should be open; accompanied by relevant caveats and health warnings to avoid misuse. Issues were raised about disclosing personal information. Whilst the foreword of the consultation clearly states that personal data should be protected, from responses received there are challenges to this in relation to requirements to disclose public data. For example, we are aware of a case where, in disclosing all spending over £500, sensitive payments to people involved in child protection were inadvertently (and briefly) released by one Council. This raises the general point that, whatever tests are applied, public service providers cannot publish everything without careful redaction if sensitive data is to be protected.

C.3 Where releasing data is not deemed to represent value for money, it is suggested that clear guidelines would need to be issued to public service providers to ensure a loophole was not created for data to be held behind. General points on this included:

- The presumption is that all data be made available.
- Value for money tests should rely on each public service providers' relevant policy and guidance (e.g. green book).
- In cases where value for money tests are not passed (a decision explained in writing to the requestor) an opportunity should be made available for the requestor to bear some of the costs.

C.4 It is recommended that where organisations are part funded by the public sector it is written into contracts and agreements as to what data needs to be made open. Guidance could be issued suggesting a minimum level of public investment in an organisation after which open data principles would apply.

C.5 To encourage or ensure publication, a simple and quick online method to make open data available is required. The South West Observatory has experience with previous interoperable online systems, which struggled to take off due to the cumbersome process of uploading / editing and categorising. Annual reporting by public service providers of datasets requested could also be considered, although this would pose an additional capacity challenge for organisations involved. One practitioner suggested "*[whatever method] something simple has to be adopted if we are to meet requirements.*"

D. Policy Challenge Questions

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?/ How would we ensure citizens have a legally enforceable right to access easily, use and re-use data held about any publicly funded service?

General Questions

D.1a These questions, differently worded on pages 22 and 23, have a different nuance. In the first, stronger rights to obtain and use data from public service providers implies the ability to get data from local authorities, the civil service, NHS, and other public sector bodies, with the implication that this is data they generate themselves. The latter opens up the possibility of legally enforceable rights to access data held about any body receiving public funds – including private operators of publicly funded services (e.g. train operators), social enterprises, charities etc. who may be awarded contracts. The scope of this consultation needs to be clear as “commercial confidentiality” is often cited to restrict the use and reporting of data. The freeing up of such data could help evaluate policy interventions by comparing the outcomes of interventions against their intended purpose.

For example, rail ticketing data is held by the Office of the Rail Regulator and can be restricted in access and use owing to commercial confidentiality. Understanding how many tickets of a certain type, for a certain route, were sold before and after a change in operator policy could help passenger groups, campaigners or other operators evaluate whether a measure to introduce an evening restriction on day returns actually relieved the evening overcrowding as per the stated intention, moved the peak to another time or simply raised more revenue from a captive market without shifting travel patterns.

D.1b These questions should be considered in the context of the policy challenge question about meaningful open data. Data is collected for a number of reasons, and sometimes - and particularly with administrative data - just because it is possible to collect it. This begs the question whether all data should be published regardless of demand, or whether all data should be made available on demand.

D.1c With rights come responsibilities and this consultation does not consider the onus on the user of data to use it appropriately, taking heed of caveats and limitations. When data is misinterpreted openly and publicly, it can be challenged openly and publicly as demonstrated by the poor use of crime statistics during the 2010 election campaign <http://bbc.in/7YIocT> . How do we guard against the poor use of data or, more positively, encourage the best use of data, to inform decisions, policy or actions by individuals or organisations who themselves are not subject to FOI and where the data assumptions may be hidden in the outcomes (whether by accident or design)?

Questions from p25 –

D.2 Legislation to require public service providers to publish all data might ensure publication of more data. However this has little bearing on the potential need for, or use of, that data. In the spirit of localism, dialogue with actual and potential users should establish what they are seeking to achieve and therefore what data (not currently available) they would like to have access to. The most useful data may in fact not exist, which could highlight gaps and the need for further research.

D.3 The Information Commissioner may be the best option to safeguard the right to access / right to data and is preferable to creating an additional role. However, the National Statistician and ONS may also have a role to play.

D.4 The right to access and a right to data have to be balanced with the right to have personal details held safely. Local safeguards are needed to ensure that access to personal information is only given to the individual concerned or authorised personnel, and that attributable data is not published. In cases where the disclosure of sensitive data has been narrowly avoided (example – C.2) this has been a result of staff vigilance rather than existing safeguards *per se*. Where there have been well publicised breaches of data protection in the past, e.g. CDs with personal details being posted/lost, they have often been down to junior members of staff – or external contractors – unaware of Data Protection issues. This highlights the need for organisations to train and regularly remind all staff of their individual responsibility, to put in place locally enforced checks and balances to ensure “existing safeguards” are properly implemented.

D.5a As local partners in Devon have stressed, whilst the principles behind open data are supported the process will not be quick, easy or cheap. Costs for this are potentially significant; the work is likely to be a back / middle office requirement, but with reductions in the public sector this could fall to frontline workers in which case resource implications would be significant. In addition, whilst it is generally accepted that when automatic transparency and publication is achieved FOI requests might tail off, there is no sign that this has yet occurred nor will it in the short-medium term.

D.5b The question of associated costs for releasing data raised comments about the evidence behind exactly who wants access to the data and what evidence there was to support the claims around the positive impact its release could have. NESTA’s ‘Make it Local’ initiative was highlighted as a potential way to help maximise public service resources and practitioners’ time, through engaging with communities to identify specific needs to pinpoint relevant data.

D.5c The scope of an enhanced right to data needs to be clarified.

D.6a In terms of ICT contracts, clear guidance or legislation was suggested to ensure Open Data standards are embedded through the incorporation of standard clauses. However it may be less an issue of ICT contracting and more about embedding a ‘default sharing’ discipline in the data producers/gatherers.

D.6b Different data is gathered in different ways for different purposes. MS Excel is on every desktop but people make their own spreadsheets to suit their own purposes and while Excel formats are highly preferable to tables embedded in PDFs, this is a long way from interoperability. Over the years there has been huge enthusiasm from

some quarters about the potential of interoperable data systems, however with data gathered for different purposes in different formats, the stumbling block has always been the amount of work needed to cleanse and standardise existing formats. Time series are important to statisticians and back data in some cases as important as current data. It may be more effective to state where data should be stored rather than how and in what format. If different organisations and private sector companies are developing parallel systems, incompatibility is likely in the short-term. However, a number of different systems already work alongside each other with informed users knowing where to go and how best to use the data available.. Nationally, NOMIS is indispensable; local information systems exist via local authority websites, many using instant Atlas; the Public Health Observatories have a wealth of data available through tools grouped for specific issues; the Homes and Communities Agency's SIGNET tool makes available a huge number of datasets through an interactive mapping tool, as examples of a few of the many data stores available via the South West Observatory's network.

D.6c In all these cases the ICT standards are less important than the data standards and the trust that can be put in the outputs. However there are challenges for the uninformed user in finding the data they require across a number of different sites. The South West Observatory has addressed this by developing a partnership network of publicly funded information providers by theme and locality in the South West of England, providing a web portal to their various sites together with a help desk function – co-ordinated by a social enterprise. We are happy to discuss how we could help enhance and roll out this sort of model nationally.

E. Setting Open Data Standards

What would standards that support an enhanced right to data among public service providers look like? /There should be a presumption of high quality publication for all data that is created with public funds and government must be held to account for meeting that standard.

General Questions

E.1 There is an inherent tension in the wording of the two questions/statements – with one giving a response to the other. "All data created with public funds" would include those created by academic institutions in receipt of research grant funding, and the data behind individual, one-off, research reports commissioned by all bodies in receipt of public funds as well as 'the public sector' as traditionally thought of. While this would be no bad thing, it could raise questions of IP with some institutions, and there would be a cost implication in making available the 'back catalogue' across the board (publication reports do not usually include all the underlying data). A sensible approach might be to set a starting date for such data to be published as a matter of course – if this were the intention. Even simply making the list of past research would have resource implications, although there could be future cost efficiencies if duplication could be avoided, methodologies repeated and data therefore gathered in a comparable way.

Questions from p28 -

E.2 Guidance developed in dialogue with those expected to comply at the operational level, together with oversight by the ICO, were suggested as means to achieve compliance on high and common standards to allow usability and interoperability.

However, it must be noted that making existing data interoperable is likely to be resource hungry, as would be setting up a new system from scratch. SWO has had some experience in this.

E.3 Consistent standards are essential for all performance indicators, and as such, if the government is to provide best practice examples to improve the provision of open data overall, consistent standards for collecting user experience across public services would be essential.

E.4 The suggestion of creating a new scheme for accreditation of information intermediaries was seen to be an unnecessary, additional layer of bureaucracy.

F. Corporate and Personal Responsibility

How would public service providers be held to account for delivering Open Data through a clear governance and leadership frameworks at political, organisational and individual level? / Public services must show leadership in making itself as open as possible.

Questions from p30 -

F.1 Responses on how to ensure public service providers make a commitment to Open Data, while respecting privacy and security considerations, suggested support and guidance needs to be provided whilst the real costs/benefits need to be kept in mind.

F.2 Given the tensions which exist between opening up data and protecting privacy, there was some agreement that board level responsibility for both matters was important. However there was a view that this responsibility needs to be cascaded and embedded in all levels of public sector staff and another view that it is neither fair nor workable to expect one individual to get the balance right each time. In terms of oversight, it was suggested that it may be better to have a separate structure for monitoring privacy issues before/when data is released.

F.3 Responses suggested support and guidance would be preferable to any sanctions framework to enforce a right to data, however sanctions were considered appropriate in breaches of the Data Protection Act.

F.4 On what other sectors would benefit from having a dedicated Sector Transparency Board, no opinions were expressed.

G. 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? / Public services should only collect and publish data that is meaningful and useful and stop collecting data that has no value.

General questions

G.1 All data is meaningless if taken out of context and without any purpose for its application. Each dataset should be understood in the context that it was collected

and with the caveats to its use put in place. The issue is not in the publication of data *per se*, it is whether data is fit for its intended purpose, and whether potential purposes are fully understood.

Questions from p31/p32 -

G.2 Responses suggested existing and known data platforms should be used to disseminate and publish information. In the case of social/demographic statistics, NOMIS web and/or the ONS website were both cited.

G.3 Value in data should be established on the basis of its relevance and usefulness to actual or planned policy outcomes and its level of usage. User testing, feedback and engagement with data users and publishers, were all suggested as ways to judge how valuable data is.

G.4 Government spending, monitoring and evidencing of outputs against business plan targets, were all cited as examples of data that should be collected and regularly published by government.

G.5a It seems impossible to make any blanket description of data collected unnecessarily. However, there should be clear, justified reasons in collecting data in line with policy considerations and value for money tests applied.

G.5b On this subject, the Localism Bill introduces policy to free Local Authorities from the burden of data collection and performance management processes. However, the Single Data List includes a number of datasets still required to be collected by Local Authorities. Some of these are onerous to collect and Local Authorities should continue to be engaged with (further to the original consultation) to ensure requirements are manageable and that data collected is used and useful.

G.6 Data quality should be defined by accuracy, timeliness and accessibility. Government releases should always be of the highest possible quality, and where revisions are required this should be clearly stated.

H. Government sets the example

In what ways could we make the internal workings of government and public sector as open as possible? / Public service providers should lead the way in open data, ensuring that internal workings and the underlying data behind advice and decisions are published.

Questions from p33-34 -

H.1 A central government portal that is accessible from all departments and includes relevant categorisation for all government departments would be useful and set a clear example about open data principles. It was suggested that any such portal should utilise existing tools, including NOMIS web, rather than reinventing the wheel. However, the data should be findable wherever it would be logical to look for it. In some cases, particularly if a report was press released by a department, or relates to departmental policy, this would be the departmental portal. Where the data actually sits in either case is immaterial as long as links actually work.

H.2 Key factors should include: relevance to key policy outcomes (actual or planned), level of usage or requests for data, and how useful data is for decision-making, monitoring and accountability.

H.3 Existing data at a more detailed level is of much greater value than a 'scatter-gun' approach to releasing new datasets. However, the objective of government should be to release *more* data with a *high degree* of detail.

I. Innovation and Open Data

To what extent is there a role for government to stimulate enterprise and market making in the use of Open Data?

Question from p35 -

I.1 Government may have a role in stimulating innovation in the use of open data. This may include seed-funding for small projects, publicising details of innovative schemes, and supporting organisations attempting to promote innovation and use of open data. The government should also promote best practice among organisations, including open data champions in local authorities. However, as the Guardian's Data Blog showcases, there are lots of examples of ideas and initiatives not stimulated by government already making use of data that is open.

J. Conclusion

J.1 The South West Observatory network is broadly in favour of the principles behind Open Data. However there are concerns about the resources required initially to meet open data aspirations against the evidence for the actual or potential demand for information.

J.2 Concerns remain regarding the publication of poor quality data; the responsibility of individuals and organisations to understand and heed caveats on data use; the issues around checks to ensure personal data is not released; and the resource implications in achieving compliance.

South West Observatory

enquiries@swo.org.uk

October 2011