Open Rights Group Consultation Response: Making Open Data Real

Table of Contents

1. Introduction and general issues

Open Rights Group broadly welcomes the consultation paper, as it makes a clear commitment to generalise proactive disclosure of data by public bodies, under an open license, complemented by a new 'Right to Data'. We welcome this opportunity to help the government maintain the momentum that has built up behind the open data agenda. It is also good news that it contains a welcome proposal for the extension of transparency policy to non-government providers of public services.

Before engaging with the specific questions asked in the consultation below, we will look at the wider issues in the policy context, and the way these proposals are being presented.

Privacy, and attempts to sidestep the issues

The consultation paper clearly states that privacy trumps transparency, but it then appears to propose that this will only apply to a small amount of data that contains directly identifiable information. We believe that privacy considerations should be extended to data that could potentially be used to identify an individual.

ORG's view is that the independent review by Dr Kieron O'Hara (O'Hara, 2011) covers the main issues and sets a good starting point for discussion. We are somewhat dismayed to hear that the independent review will not be given any special consideration and will be treated as simply another submission to the consultation process.

We share the concerns, covering several of the issues, that are raised by O'Hara's review:

- Bad record of handling personal data in the public sector
- Weak implementation of EU data protection in UK
- Lack of technical expertise, including within the regulator the Information Commissioner's Office
- Risk of re-identification of anonymous data (very real but difficult to quantify)
- Mixing public data with social media, private and other data

Until now the short-term priority for the open data community has been avoiding privacy being used to stall the momentum for openness and transparency. However, we are very concerned about the way widespread sharing of "anonymised" personal information is being proposed in the document and also in some consultation related events. Many people have told us they share these concerns.

The consultation seems to confuse different ways of using and releasing data. It presents a confusing mix of open data, 'big data' and the controlled sharing of data. This is not optimal for a proper analysis of the policy, and when it comes to

privacy the implications are completely different. Sharing an individual's medical data between government departments, or even with vetted researchers is not the same as putting it online as linked data, giving up any control over it.

A critical aspect here is the reliability of anonymisation technologies. The field is under heavy development[1], and it would be too risky to take blanket measures opening up whole data areas. Anonymised data cannot be considered secure. We would propose a slower process where anonymisation technologies are openly peer reviewed before implementation to have an opportunity to spot any weaknesses. Once implemented there should be a responsible disclosure mechanism for researchers to inform of security concerns without going public. Other than this, the process should apply recursive transparency in the choice and assessment of anonymisation technologies.

An added concern is that we cannot see privacy in relation to open government data in isolation from the wider trends for data commercialisation and population profiling. There are emerging issues around power imbalances from information asymmetries and democratic control of big data that do not fall under a narrower remit of privacy or individual data protection.

In this sense, proposals in the consultation for personal service use data to be provided to individuals should be cautiously examined. It is quite reasonable to assume that a market will develop with companies offering analytical services – for example health -- in exchange for access to personal data. The terms and conditions of these services could easily allow for individual health records to end up aggregated in huge privately run databases. Thus, following the example in health services, insurance companies or prospective employers could end up having access to the full medical records that are now inaccessible. At the very least we can imagine the development of a private health reference system, along the lines of credit reference, possibly run by the same companies.

Choice and open public services

The consultation makes clear that open data is designed to support the Open Public Services White Paper, which introduces the principle that "any willing and qualified provider" -- state, business or non-profit -- should be able to deliver public services.

Our view is that these two approaches for improving public services – public accountability with open data and marketisation -- should each be assessed and implemented on its own merits, and not be introduced as a package. Citizens should always have access to information about performance and costs of public services, but the mechanism for intervention could be other than market choice.

We do not wish to engage with the merits of marketisation and the Open Public Services agenda. Our concern is the risk of generating hostility to open data if it is perceived by sectors of society as simply an instrument for privatisation. Wider engagement from civil society in the transparency agenda could help lower this risk, but a clearer separation of Open Data and Open Public Services would be important.

[1] We provide a more detailed assessment in the consultation responses below.

The focus on market choice takes open data from citizens' engagement to consumption of services. Many open data advocates would argue that political participation should be widened, not narrowed to one aspect. In this respect the absence in the consultation of a real examination of the potential of open data to contribute to a more engaged and cohesive society is a little disappointing.

Quality and outcomes

It is widely accepted that benchmarking increases quality. What is less clear is how public scrutiny provided by open data in and by itself can improve the quality of the data or have a direct effect, without additional measures such as increasing users' data 'literacy'. This is a critical aspect of the policy. Below we look at some reviews of the evidence of the use of information in public services that show a more complex picture.

Evidence of information use on Healthcare choice

- *Competition between healthcare suppliers and patient choice are not necessarily the same.*
- *Published data has only a small impact on consumers because decision-making in health relies on specialist professional advice.*
- *Information in health care markets is often too complex for direct use by consumers. Outcomes, particularly quality, are very difficult to measure in health care, and choices are often made on dimensions that receive large amounts of public coverage, such as hospital 'superbugs' and cleanliness rates rather than on clinical measures of quality*
- *Purchasers use information on providers to a greater extent than patients.*
- *Data is most widely used by providers themselves and they appear to respond quickly to the incentives given by the information. However, this can lead to improvement on the published criteria, which is not necessarily the same thing as improving actual outcomes.*
- *Need to provide full picture to avoid "gaming". Examples of manipulation of the data from the UK include the re-categorisation of patients during the 1990s to reduce published inpatient waiting lists.*

(Propper, 2010)

Evidence of information use on schools choice

Parents may care about teaching quality but they cannot observe this directly. This means that the design of school accountability measures (Wilson 2010) is critical to how parents interpret quality. Whatever factor is chosen to be measured will become

a focus for parental interest, and so prompt a reaction from schools. Balancing the need to produce measures that are easy to understand with the desire that they reflect the quality of teaching that takes place is not straightforward. The Contextual Value Added measures of school quality that are published in schools performance tables are an admirable attempt to indicate the quality of the school, rather than the quality of the pupil intake, but surveys suggest that parents still focus on the proportion of pupils gaining 5 or more A to C at GCSE because it is a metric they can understand.*

What is unclear is the extent to which the development of better quality metrics could reduce the extent to which parents favour schools with affluent intakes, sustaining the relationship between pupil characteristics and school popularity.

(Allen & Burgess, 2010)

The consultation appears to take for granted that more informed choice powered by open data will automatically deliver better outcomes, but if we look at the available research on the use of information for exercising choice in public services, this is not clear.

Effective use of data by citizens will require both a minimum of data literacy and intermediaries. There is a risk that market consolidation will reduce the diversity of analysis and data intermediation.

Capacity of public sector and civic collaboration

There is a need to build capacity to analyse and promote open data at every level, but in the context of cuts it is hard to see how this will happen. The skill base is currently in the private sector – particularly for high end, complex analytical skills. Whilst there are 'civic hackers', and there is some overlap, there needs to be investment in the skills base to make sure capacity is deepened within the public sector and civil society.

However, civic hackers, or even consultancies such as Dr Fosters, are not going to run public hospitals. There needs to be some process in connecting data experts, public sector workers and service users in order to contribute to public service improvement. The proposed mechanisms of customer feedback and choice are not enough in our view. In short, there should be the coordination and skills to help users and public sector workers ensure that open data is working towards the public interest.

We believe that there should be civil society representation – beyond civic hackers and technical experts -- in decision making teams/panels and in review panels at all levels that decide what data is going to be released and under what licence.

2. Responses to consultation questions

An enhanced Right to Data

Definitions of key terms

The strict definition of data in the consultation may be too restrictive. The definition should not exclude data that has been subjected to any analysis or processing.

The consultation appears to have an almost exclusive focus on external flows of data, but very little on how public bodies can improve their internal processes through open data.

Range of organisations covered by the proposals

The FOIA amendments to cover private bodies delivering public services are welcome. This should cover not only performance data of the service delivered, but the financial and organisational data required to properly assess whether the organisation is providing value for money.

There is a longer term perspective too. Transparent government cannot exist without more transparent societies. Private bodies -- particularly charities, publicly traded companies and those working for government – should expect increasing demands for transparency in their operations. These should initially cover environmental, ethical and financial aspects, including tax transparency.

The description of right to data in this document as applying to "public service providers" could be interpreted in a restrictive manner. Governments perform a variety of functions in running the country, including state tasks that do not fall under public services, such as war and negotiations of international treaties, and it is important that transparency and openness are not just focused on frontline services at the expense of a more holistic view of open government.

One final aspect that should be clarified is the extent to which Whitehall will work with the devolved administrations to create a UK-wide open data policy. The is currently unclear.

Stronger presumption for publication

We agree that an "open by default" policy should guide the release of data. We understand that this may take some time, but making the implementation of the policy contingent upon the renewal of ICT systems - as it appears to be proposed - could be problematic.

The UK has a very patchy record of public ICT delivery, and we believe it would be beneficial to produce an independent timetable for the expected phasing in of the new policy across government bodies.

The policy must make very clear that data/information published by default is also reusable by default through the Open Government License. The UK country action plan presented to the Open Government Partnership ("United Kingdom | Open Government Partnership," 2011) says "public by default", instead of open. This is a major difference, and we hope that the "open and reusable" option will remain the main policy.

Generally, the already agreed Public Data Principles should serve as a guide. We find it confusing that they are not directly referred to in the consultation document.

Requestors paying for data

Current fixed fee charges for access to personal data act as deterrent and may not bear any relation to the real costs involved. Therefore they should be removed.

The principle that cost should not be an obstacle to transparency is good, and allowing requestors to pay the extra amount above the statutory limit, instead of being rejected is potentially positive. The basic principle should be that a genuine extra burden would be placed on an organisation. There are several aspects that must be ensured:

- The body must make a proper case, and the costs involved in rejecting the publication of information must not be higher than the cost of releasing it.
- There should be some standard costing guidelines based on types of activities and volumes to avoid abuse.
- Data / Information released to one citizen or organisation prepared to pay extra must be disclosed to the public at large as well, in a suitable manner.
- Test for a public interest in the information being released that forces the body to cover the costs of the additional burden. This should also force public bodies to incorporate the release of that data in future business planning.
- Affordable and fast appeal process with strict time limits, via either the ICO or a new Data Ombudsman.

Data ombudsman

There are two separate issues in regards to the framework required for ensuring that the policy is properly implemented: ensuring compliance and providing an adequate governance.

ORG agrees that there is a need for an external body to ensure compliance with proactive data publishing including formats. Ideally, the Information

Commissioner Office (ICO) should carry this function, as the new right to data is within their FOIA responsibilities.

However, there are serious issues with capacity and the lack of technical skills in the ICO that may make this difficult in the short term. The ICO is currently in no position to advise on suitable data formats. The Office for Public Sector Information could have the capacity to provide advice, but has not proper powers to enforce public bodies' compliance with an amended FOIA.

In the short term a bridging body – Data Ombudsman – could be formed that had access to the licensing and technical skills of OPSI, while relaying cases already prepared directly to the ICO when the use of statutory powers is required.

We generally agree with the proposed new powers for the ICO to:

- carry out non-consensual audits
- administer disincentives for non-compliance (including organisations)
- decide a public interest test for publication of data with cost override

In the medium term the function, powers and capacity of the ICO must come under deep reevaluation, as part of the review of the EU Data Protection framework and its extremely unsatisfactory UK implementation. This is such a long standing and important issue that we would prefer to engage with discussing any enhanced powers for the ICO elsewhere, as part of the wider reforms required.

For example, the redress for privacy violations is very limited, and in the context of opening data, this should also be considered.

Are existing safeguards for privacy and data protection enough?

The consultation makes clear that private data will be out of the scope of the proposals. The question then is what data should be considered private. From a privacy perspective, if data is personal data, as set out in the Data Protection Directive – that is, if it is data that can be linked to an individual – then the assumption should be that this data is private. That is, the default position should be one of privacy. There is some personal data that is public – for example directorships in companies – but that data needs to be specifically set out and agreed.

'Anonymised' data

The limitations of 'anonymisation' processes need to be acknowledged and their implications understood. There is evidence to suggest that much supposedly 'anonymised' data can be 'de-anonymised', by combining it with other, often public, data sources. In 1997 Latanya Sweeney demonstrated that by combining an anonymised hospital discharge database with public voting records a range of

identifiable health data could be produced.[2] Computer scientists have continued to work on de-anonymisation – their models are getting substantially stronger and more applicable to the kind of data now being generated on the internet.

In a 2008 paper, Narayanan and Shmatikov of the University of Texas demonstrated by combining the databases of Netflix and the online movie database IMDB that if you knew the county someone lived in and one movie that they had rented in the last three years, they could be uniquely identified 84% of the time. Moreover, they suggested that their results could be generalised – and applied to most other similar databases.[3] The implications of the continuing work in this field are significant. It can be argued that anonymisation is to a great extent illusionary[4] – and even at best it means that it needs to be considered very carefully and its weaknesses taken seriously.[5]

As far as the Making Open Data Real agenda is concerned, this means that 'anonymised' data such as medical data should not be included. If, as the evidence suggests, this data could subsequently be de-anonymised, it would put some of the most sensitive and personal data possible into the public domain, or, perhaps even more damagingly, make it available to be bought and sold. That, above all, should be avoided.

Aggregated data

If the idea of including medical data is considered important, then the data should be aggregated rather than anonymised: that is, rather than individual anonymised records, overall summarised statistical information could be considered. The aggregation process, however, needs to be done very carefully, to remove any prospect of disaggregation.

What applies to medical data would apply equally to other 'anonymised' and aggregated data – survey results, census returns and so forth. Anything from which individual records can be taken and potentially de-anonymised needs treatment in the same way. It is important also to understand that it is not only the most obviously 'sensitive' data that should be protected – less sensitive data can be used to derive sensitive information. If we protect the records of who has

[2] SWEENEY, L. 1997. Weaving technology and policy together to maintain confidentiality. *Journal of Law, Medicine and Ethics*, 2**5**, 98-110.
[3] NARAYANAN, A. & SHMATIKOV, V. 2008. Robust De-anonymization of Large Sparse Datasets. *IEEE Symposium on Security and Privacy.* 2008 ed.Available online at http://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf
[4] As suggested, for example, by Michael Colao at a meeting of the Society for Computers and Law in March 2011. See http://www.scl.org/site.aspx?i=ne19845
[5] See for example the work of Paul Ohm, in OHM, P. 2010. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review*, 5**7**, 1701-1778. Ohm analyses the work of computer scientists from Sweeney onwards and suggests that a full understanding of the weaknesses of the anonymisation process is required if methods to protect privacy are to be effective.

diabetes but do not protect the records of who buys 'sugar-free chocolate for diabetics', we misunderstand the way that data can be used.

Summary

- Private data should not be made public, and hence be part of the remit of the Making Open Data Real proposals.
-
- If data can be linked to an individual, it should be assumed to be private unless there is a specific reason to make it public
-
- Anonymisation processes are far less effective than they may appear.
-
- 'Mundane' data can be used to derive sensitive data
-
- The best approach would be simply to exclude from release any data that can be broken down into individual records, whether identified or anonymised.

Setting transparency standards that support an enhanced right to data

High and common standards

ORG supports the use of open standards as a general principle, but the setting of common standards should be a participative process with sufficient flexibility, rather than a one-off top-down exercise. These standards will in all likelihood be domain specific and therefore the Sector Boards will have an important role. The current proposals by OFGEM for the simplification of the energy market show the importance of defining common indicators to avoid confusion by end users.

Licenses

The proposed development of additional licenses for public bodies based on OGL is of course welcome. However, there have been suggestions that some of these licenses could include non-commercial restrictions. While we do not wish to take an absolute stance, we must point out that non-commercial clauses make almost impossible the development of the open data ecosystems required for the expected benefits of the policy to be realised. Non-commercial clauses are difficult to enforce and generally create a two-tier market where the same information is relicensed fully commercially elsewhere.

While this may be understandable for individual artists, it makes no sense in an open data environment. Research from Canada on the use of another type of restrictive license - share-alike - by municipalities evidences that they bring more problems that advantages in the context of public data (Fewer & Mewhort, 2011).

Accreditation Schemes

It is not clear whether the proposals on the table refer to an accreditation scheme for data intermediaries, public bodies, or both.

ORG would not be against a Right to Data Assurance Framework for public bodies. We assume this would work similarly to the Information Fair Trader Scheme, run by OPSI[6]. However, positive incentive measures such as kite marks should not be a substitute for clear enforcement of Right to Data as discussed in the previous section, and ultimate ministerial responsibility on ensuring releases for each department.

We find more problematic the idea of an accreditation scheme for data intermediaries. This would have some attractive aspects in ensuring some quality, but overall we believe the needs of citizens would be best served in ensuring diversity of intermediaries by removing barriers to the participation of small businesses and civic groups. Such a scheme would risk creating a two-tier market and could act as a barrier to organisations with less resources. We can question, for example, whether MySociety would have been able to develop had an accreditation scheme been in place.

Open Government Partnership

The proposal to channel the UK's international efforts in promoting open government data through the Open Government Partnership should be treated with caution.

We support the direction of the general commitments included in the OGP declaration of principles, specifically around increasing access to information, civic participation, fighting corruption and promoting access to new technologies. However, each country voluntarily provides their detailed actual commitments. While we can see how this ensures flexibility, it may be allow some countries to gain the same "branding" of openness with little actual delivery.

The ability to "pick and mix" areas for improvement may make it difficult to compare countries. In the case of UK, the commitments are centered around improving public services, but other countries focus on natural resources, corruption or civic participation.

The obligations regarding civil society consultations in the OGP are very welcome, but we are concerned that in the initial stages of preparing the UK involvement and national commitments, the extent of consultation with civil society appears somewhat limited.
Calls by NGOs to use the OGP to improve international transparency over extraction of resources ("Open letter from PWYP UK to Prime Minister David Cameron on OPG country action plan," n d) were not in the UK action plan. We would expect the development of a consultation framework with civil society to

[6] http://www.nationalarchives.gov.uk/information-management/ifts.htm

take place at the earliest.

Corporate and personal responsibility: governance and leadership framework

Ensure day to day decision- making commitment to Open Data, while respecting privacy and security.

The move to open data requires a fundamental cultural transformation of the public sector. This can only be achieved by integrating openness at every level in job specifications, departmental business plans, personal development plans, inducements and so on. The use of data to monitor and improve processes is a fundamental requirement of any modern organisation, so this cannot be fully isolated from day to day operations.

Line management structures are already in place, and there is no need to create a completely separate system for ensuring open data is delivered. However, formalising an Information Governance Group would be a positive step. The following example from the health sector may be excessively complex for other areas but gives a good indication of how wide are the implications of data and information:

"(…) information governance group which comprises of the director of information, director of operations, head of human resources, head of IT services, clinical information manager, IM&T manager, risk manager, legal claims manager, data protection coordinator, EPR clinical lead, assistant facilities manager, Caldicott implementation project manager, IHCS manager, and the communications manager. It meets for two hours every two months. It is a coordinating group, which monitors trust-wide security matters in its broadest sense and reports to the trust board and to the necessary heads of department if there appears to be problems."
(Roch-Berry, 2003)

Personal responsibility at Board-level

We must partly distinguish the expertise required to implement the new agenda and the responsibility for it. Every public sector body should have someone at the board level that can understand and help deliver the open data, including both the technical and legal aspects. This should not be reduced to an evangelist role, but one with executive capacity.

Many public bodies already have a non-senior person responsible for access to information and general FOIA responsibilities. However, it is unlikely that these officials have the required technical expertise, and the same will probably happen when the open data role is created at board level. This lack of integration of skills may be one of the key obstacles to the delivery of the policy, and an area where investment will be required.

It is however impossible for most public bodies to have the in-house expertise to deal with increasingly complex anonymisation technologies. This means this expertise will have to be created and shared.

The health service has a very sophisticated system of roles and procedures for dealing with patient confidentiality, and one of the proposals in the consultation is extending the so called Caldicott guardians to other public bodies. Caldicott guardians (The Caldicott Committee, 1997) have a very specific mandate to protect patient privacy, and it would be good to see the same level of commitment in other public delivery areas that until now have not been subjected to the same level of scrutiny.

As open data is implemented there will be increasing conflicts over data control in other areas, but not necessarily over individual privacy. For example, the release of Police crime maps had implications for both individuals and whole populations in particular areas. This means the role of protecting service users' data will need to adapt to the circumstances.

We do **not** believe that the same person should be responsible for open data and protecting the privacy and confidentiality of public service users. These two roles have different priorities, and if taken by one individual it is unavoidable that they will err on the side of caution towards one or another. Therefore, it must be the responsibility of the head of the organisation to ensure that both openness and privacy are protected. In most public organisations, the ultimate responsibility for serious or systemic failures of any kind falls on the Secretary of State for the relevant department. Open data should be no different.

The individuals responsible for privacy and open data will have to work very closely with each other and others in the organisation. A local Information Governance Group would be the best space for this to take place before it reaches the board.

Sector Transparency Boards

We welcome the creation of transparency boards for each sector because particular domains require a level of expertise. They will be a positive addition provided they fulfill the following conditions:

- Meaningful participation of civil society, including: service users, unions, advocacy groups, civic hackers, etc.
- Include a privacy expert / Caldicott guardian.
- Include the technical expertise to understand both open data and privacy enhancing technologies.
- Have enough power to influence the agenda.
- Are integrated in a wider program for delivery of open data.

The Sector Boards are not a substitute for a national transparency board. One of our key aspirations is the widening of the Transparency Board along the lines presented above, and giving it a clearer mandate. Representatives of the civil society consultation framework required by the OGP should participate in the

board.

Meaningful Open Data

We believe that a demand approach to data, with fluid communications from external users of data and swift action to release, can be more beneficial that focusing on the supply side.

Creating comprehensive and up-to-date catalogues, inventories and Information Assets Registers has not been a successful strategy so far, and it is unclear how this can be changed without substantial resources and ICT changes.

This is not to say inventories should be completely abandoned. Catalogues can be particularly useful to track new updates and cross searches for data.

Data quality

On the supply side, resourcing the creation of adequate metadata and documentation at the point of origin could be the single most effective intervention, rather than looking for "quality". Contemporary understandings of "quality" centre on customers' needs and expectations, not on intrinsic properties of the product, but applied to public data we can see a clear role here for intermediaries to deal with different "customers".

An iterative improvement process in dialogue with users is different from polishing the data for its own sake. Further "polishing" of the data may suffer from "gold plating" and diminishing returns. A positive step would be to allow data users to feed back improvements to the data in a controlled manner.

Data value

The proposals to prioritise data for inclusion in inventories are very problematic, as any definition of value is bound to be different depending on the viewpoint.

Data that is important for internal operations may not be the same as data for accountability purposes, and data with high economic value to businesses may be different altogether.

Our concern is that any attempts to create a single measure of value will end up dominated by the most powerful interests. Therefore, if a form of prioritisation cannot be avoided it is important that this is carried out with a diverse input. The Sector Boards should provide general guidance but may not be able to deal with the level of detail required for individual datasets.

Unnecessary data?

Further proposals to stop the collection of certain data should be treated with even more caution for obvious reasons.

Proposals circulated in consultation meetings to "draw on lessons from the Government Data Review [7] to focus resources on the collection of essential data" need clarification. Besides, while that document contains some sensible recommendations around data handling processes and powers for the ICO, others are far less clear. For example, the recommendation to give the Secretary of State authority to remove barriers to data sharing in legislation is quite dangerous. The review may also be outdated in relation to anonymised data.

Government sets the example: open the internal workings *of government and the public sector*

The proposals in the consultation paper for opening up the workings of government by publishing research and surveys underlying public policy are very welcome. However, in the consultation meetings and further discussion with other parties we find that the focus appears to have shifted to introducing increased data sharing and facilitating public data to external researchers. This may be a positive step depending on the circumstances, but it should have been made clearer in the original documents.

The concrete proposals circulated in consultation workshops to implement "ESRC recommendations for improving use of data for research and policy" are unclear. One important issue here is that the examples presented throughout the consultation generally confuse open data, big data and controlled sharing of data. Researchers normally work in controlled environments, and would have access to data that would never be published as open data. Further to this consultation, it may be important to clarify these aspects. At present it appears that changes to data policy involving increased data sharing, but not strictly open data, are going to be introduced as part of the package, and we believe these should be consulted separately. The privacy implications are quite different in each case.

This is illustrated in another concrete proposal presented in a consultation workshop for: *"Research Councils and Cabinet Office Taskforce to link anonymised data across public services for researchers and liaise with Sector Boards. Promote proactive sharing of anonymous data outside traditional Whitehall silos"*. As mentioned elsewhere, there are serious doubts about the safety of anonymised data, and there are serious differences between providing a better system for researchers and generalised proactive sharing. We believe this proposal should be put on hold until the privacy implications are properly understood.

Innovation with Open Data: should government stimulate?

In line with most other open data initiatives the current policy drive expects that a market for information and ancillary services will flourish. This policy goes further however, with open data is presented as one of the pillars of the

[7] We can only assume it refers to this document http://www.justice.gov.uk/reviews/datasharing-intro.htm

forthcoming Strategy for Growth. Our view is that it is unclear whether such high economic benefits will accrue in the short term, but we must ensure this doesn't detract from the commitment to open data.

There is growing concern among open data advocates that the economic value of public data in the short term may be inflated. The McKinsey report (Manyika, 2011) quoted in the consultation claims the value of better data handling to the European public sector is €250 billion. Assuming a relationship of this value to the countries' proportion of EU GDP (Wikipedia data), the UK public sector could make just over £30 billion in:

- Operational efficiency savings.
- Reduction of cost of fraud and errors.
- Increase in tax revenue collection.

We have to ask ourselves whether this is realistic. Besides, how much of this relates to data being open? As mentioned above, both McKinsey and the UK government conflate: Big Data, Data sharing and open data. For example, both car manufacturers and state health services could improve their internal processes by sharing and digesting big data without making it public. They could also make data public for crowdsourcing particular things, but not allow reuse, thus not being open.

The McKinsey report has good pointers at the innovative disruption that should generate the new wave of growth, particularly damaging to businesses based on information asymmetries. However, incumbents in UK will be more protected because of the selective opening of data in the Public Data Corporation.

We partly agree with a recently published Demos paper (Leadbeater, 2011): the big data promise will only be delivered if government opens up the data and brings an innovation strategy. But sustainable innovation is harder. From the original Apps for Democracy 2008 in USA most are not working. We may need an industrial strategy with investment and wider reforms (e.g. Hargreaves review of copyright).

In any case, Open Data must be primarily driven by core democratic values, rather than calculations of efficiency or economic growth, as the latter may not materialize and the whole agenda could suffer a backlash.

Another issue of concern is the socially just distribution of benefits from access to public information. Besides, the potential detrimental effects of this policy for some groups or individuals are not yet well understood. Several of the growth areas the paper highlights: life sciences, population data mining and risk profiling are likely to lead to conflicts if data subjects suffer from increased insurance premiums, or reduced services based on behavioural choices. The impacts assessment for this policy should be broad in its scope.

For further information or clarifications please contact:

Javier Ruiz
javier@openrightsgroup.org
+447877911412

REFERENCES

Allen, R., & Burgess, S. (2010). The future of competition and accountability in education. London.

Fewer, D., & Mewhort, K. (2011). *Analysis of share-alike obligations in municipal open data licenses* (pp. 1-6). Ottawa: Samuelson-Glushko Canadian Internet Policy and Public Interest Clinic.

Leadbeater, C. (2011). The Civic Long Tail. London: Demos.

Manyika, J. (2011). *Big data : The next frontier for innovation , competition , and productivity*. *McKinsey Global Institute* (p. 156). McKinsey Global Institute. Retrieved from http://www.mckinsey.com/mgi/publications/big_data/pdfs/MGI_big_data_full_report.pdf

Open letter from PWYP UK to Prime Minister David Cameron on OPG country action plan. (n.d.). . Retrieved October 15, 2011, from http://www.publishwhatyoupay.org/resources/open-letter-pwyp-uk-prime-minister-david-cameron-opg-country-action-plan

O'Hara, K. (2011). *Transparent Government, Not Transparent Citizens: A Report on Privacy and Transparency for the Cabinet Office*. London, UK. Retrieved from http://eprints.ecs.soton.ac.uk/22769/

Propper, C. (2010). *The operation of choice and competition in healthcare: A review of the evidence*. London.

Roch-Berry, C. (2003). What is a Caldicott guardian? *Postgraduate Medical Journal*, *79*(935), 516-518. doi:10.1136/pmj.79.935.516

The Caldicott Committee. (1997). *Report on the Review of Patient-Identifiable Information*. *Health (San Francisco)*.

United Kingdom | Open Government Partnership. (2011). . Retrieved October 14, 2011, from http://www.opengovpartnership.org/countries/united-kingdom