



Open Public Services Network: response to Making Open Data Real

Intro

We welcome this consultation into open data and transparency. The government has set out an ambitious schedule and a wide range of activities, all of which are important. Our comments are intended to help identify the areas of greatest potential benefit – and therefore greatest priority. We also address the specific problems presented by these areas of activity.

Our key recommendations are

1. Prioritise improved access to large non-personal datasets routinely collected by public services. These have the greatest potential to unlock value in the public sector, empower the public and promote growth.
2. Ensure that the public have access to information at a level of granularity relevant to the decisions that they are expected use it for.
3. Divert resources away from creation of “data tables”, “polishing data” and analysis of large routine data sets towards improvements in access to underlying data sets and improvements in data quality.
4. Recognise the need, with many of these data sets, for licensing and legal arrangements that go beyond the open government licence.
5. Set a timetable for release of these key data sets (our list is given below)
6. Take steps to ensure greater access to these data sets also promotes social growth.
7. Put in place arrangement to make available all public services survey data through data.gov including all historical results. Key surveys are listed below.

About OPSN

The Open Public Services Network (OPSN) is a new not-for-profit organisation dedicated to making data about public services informative and useful to the general public. Open Public Services Network will:

1. Develop applications that use data to support more informed use of public services
2. Develop applications that strengthen public accountability around performance of services
3. Lobby for public services to provide information that meets the needs of service users

Government policy to improve transparency in public services is leading to ever greater amounts of information being made available about public services. Use of information by the public to inform decisions or hold organisations to account has been limited. OPSN exists to encourage availability of relevant actionable data and applications that engage the consumer. OPSN is focussed on expanding the market for use of data and targeting areas that are not well served by the market.

For further information please contact:

Roger Taylor
Open Public Services Network at the Royal Society of Arts
8 John Adam Street
London WC2N 6EZ

Email: OpenPublicServices@gmail.com
Tel: 07780 992 835

Rationale for recommendations

1. Why prioritise routine data sets?

These are the data sets that have most to offer in terms of the six objectives set out in the consultation. The consultation identifies three areas for open data:

- Large non-personal datasets collected routinely
- Individual access to their own data
- Better user feedback

The table below rates these activities against the six objectives (grouped into two sets of three).

Data types	Productivity, economic growth, quality and outcomes	Accountability, choice and social growth	Priority
Large non-personal datasets collected routinely by public services e.g. Hospital Episode Statistics	These datasets have very high and immediate value in identifying productivity failings in public services. The economic growth potential from developing industrial expertise in these areas is high in the short to medium term.	Understanding where outcomes are being achieved most successfully must underpin the public debate around choice, accountability and social growth.	1
Right of individuals to access and control their own service user records e.g. a school or personal health record	This will have relatively little immediate impact on quality and outcomes or productivity for most people. However, in the long run personal ownership of data is a vital mechanism to build public trust in the use of information. In the longer term it is likely also to stimulate economic activity.	For the long term development of social growth this is essential. But it will take time.	2
User feedback on service e.g. comments and suggestions from users of health, personal or social care services.	This can stimulate improvements in quality and build a “customer” focus in public services that tend to be inward looking. It is less effective in terms of productivity and economic growth.	This can have a rapid impact in terms of choice and social growth. However it needs to be balanced by data on real outcomes.	2

2. Why provide greater granularity?

The belief that great choice for the public in their use of health services, education and other public services will drive improved quality, satisfaction and efficiency has been held by the last three governments. Where this approach has been given real opportunity, the evidence suggests that it is effective. However, as a strategy, it has been held back by an imbalance in the available information – you want to choose where to have your surgery, but you only know the average performance of the hospital. If the public are to be offered meaningful choice of GP or individual hospital doctor, outcomes and satisfaction ratings need to be available at this level, not just at GP Practice or hospital level.

Creating data tables within government agencies at ever more granular levels risks using up large amounts of resource while failing to meet public needs. An alternative strategy is to allow access to the underlying event level data (see below). This will promote economic growth and is more likely to result in public information needs being met effectively.

3. Which routine data sets?

The routine data sets of greatest value are those that can be used to identify variation in outcome from public services and relate that to the resources and organisations involved in delivering these outcomes. Those data sets are:

Health:	ONS data on deaths, births and population
	Hospital Episode Statistics
	Cancer Registry Data
	National Clinical Audit data sets
Education:	PLASC data
	Higher Education data (HESA)
Employment:	Incapacity benefits data
	Job seekers allowance data
	Carers allowance data
Policing:	Administrative data on crime reporting and clear up

There are two areas where national structure routine data does not currently exist but where there would be considerable benefit in creating such data. These are social care and primary care.

4. Why make the underlying data available?

With most of these data sets arrangements to access the underlying data are weak or non-existent. In some cases there are legal obstacles to access (e.g. ONS deaths data). In others there are concerns about legality or privacy issues.

Making underlying data accessible is essential to get the full value out of the data. With all of the above identified data sets, there is a virtually infinite list of questions which the data can answer. For example from the ONS deaths data

- How many people died of road traffic accidents in the postcode areas around my estate?
- For what conditions do people most often die in nursing homes, hospitals or at home?
- How have causes of death among 16-25 yr olds in South London changed over the last 15 years?

Currently a great deal of time is spent creating websites that release these data in a range of different tables. These tables in the main fail to answer the specific questions of service users or those providing services. There is no prospect of ever being able to produce sufficient data tables to meet the information needs of the public and public sector organisations. This needs to be recognised and a new approach adopted which scales down this activity and instead focuses on release of underlying data to organisations and individuals. Release of underlying data is far less costly than the current approach of preparing data and building websites to view certain cuts of the data.

5. Going beyond the open government licence

With all the above data sets, at the most granular level, there may be risks of re-identification if data is misused. For that reason, there are circumstances where data release needs to be controlled. Those applying for data need to:

- Demonstrate secure storage of data
- Sign up to standards in terms of how data is processed and used

The aim must be to set safe standards. Where these standards are met, data should be provided. Data release must not be conditional on what the organisation proposes to do with the data – i.e. the types of information it intends to create from the data. However, data release may be conditional on ensuring organisations behave in a manner that promotes social growth and the public benefit.

6. Ensure greater access to these data sets also promotes social growth

These data sets have enormous potential to enable identification of more efficient and effective ways to deliver public services. This delivers benefits in terms of:

- More efficient public services
- Economic growth through the development of an industry to support understanding of big data in the public sector
- A new relationship between citizen public services through social growth, accountability and choice

The significant benefit will accrue to the public through more efficient public services. Some benefit will also accrue to private sector industries that help achieve this. However, it is essential that benefit also accrues to the public through social growth, accountability and choice. This can be achieved by:

- Ensuring that all organisations use public sector data commit to being themselves transparent and publishing what they do with the data.
- Supporting organisations that wish to use data to promote accountability, choice and social growth such as

- Organisations representing users of public services e.g. patient organisations, organisations supporting victims of crime etc
- Public sector organisations that wish to use data to be more transparent in their dealings with their users.
- Private sector organisations that publish information for free

7. Publish survey data

There is an enormous amount of data from surveys of users of public services which is not published. This includes a number of national surveys e.g.

- National NHS patient survey
- National NHS staff survey
- Housing satisfaction surveys
- Surveys of people in receipt of social care
- Surveys of tenants

These surveys are sometimes published, sometimes not – sometimes in complete form and sometimes only in part. Often, they are in theory available through the UK Data Archive however this organisation limits access to the public, rather than promotes it.

These surveys should be made available online with immediate effect as downloadable cvs files, by organisation and question ,detailing the number of respondents and answers for each question. This should include the complete time series of past surveys to allow understanding of trends.

In addition a general obligation on public sector organisations conducting surveys of the public to make the data from these surveys available in this format should be established.

Below we give answers to the specific consultation questions where appropriate

Consultation questions

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?

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

Through the following steps:

1. Establishing a right to data in law.

This would have the benefit of:

- i. Getting rid of existing legal barriers that exist with regard to certain data sets e.g. ONS deaths data
- ii. Clarifying the legal issues around processing of anonymised data where there remains a risk of re-identification
- iii. Establish a mechanism of redress for organisations that wish to access data but prevented from doing so

2. Establish an obligation on organisations to publish data in a timely fashion.

This is necessary in addition to establishing a right to data to embed a pro-active approach to publication. Meeting external demands for data should be the priority – since these are the data sets that are most likely to be used for public benefit. However, obligations to publish data as a matter of course are also needed to create a culture of transparency and openness.

The most valuable data sets to which business and the public need access are those where risks of re-identification exist. These data sets need an appropriate legal framework and licence terms put in place over and above those set out in the open government licence. Currently, legislation imposes specific legal restraints on particular data sets. A general legal framework for handling all such data sets should be created. This should give powers to the SoS or such agencies as he/she appoint to release data under licence to organisations meeting set criteria. It should also give rights to organisations and individuals to apply for such data and rights to independent review of these decisions.

We would recommend that within such a framework the Cabinet Office should

1. Identify key data sets that need to be available in anonymised format for release under appropriate license. These datasets are listed above.
2. Create standard license terms and thresholds that need to be met for access to each data set.
3. Set a target date for achieving this with each of these data sets.

Q1.2 Is providing an independent body, such as the Information Commissioner, with enhanced powers and scope the most effective option for safeguarding a right to access and a right to data?

Yes see above.

Q1.3 Are existing safeguards to protect personal data and privacy measures adequate to regulate the Open Data agenda.

No. The principle issue that needs addressing is the handling of anonymised data sets where there remains a small risk of re-identification.

Q1.4 What might the resource implications of an enhanced right to data be for those bodies within its scope? How do we ensure that any additional burden is proportionate to this aim?

The aim must be to release data in as close to its original form as possible. This takes less resource than the way in which data is currently released. Most data is currently released in the form of pre-prepared data tables. This takes a large amount of resource in the following activities:

- Polishing data prior to release
- Creating data tables that remove small numbers
- Creating web interfaces that allow access to these data tables

This process creates little or no value since

- Polishing of data removes information
- The data tables almost never produces the particular data table that the user needs for their specific purposes

These activities should be scaled back to the minimum necessary to support specific government requirements. Resource should then be re-focussed on

- Driving up data standards at source collection
- Meeting the needs of data users

Q1.5 How will we ensure that Open Data standards are embedded in new ICT contracts?

No comment.

2 What would standards that support and enhanced right to data among public service providers look like?

Q2.1 What is the best way to achieve compliance on high and common standards to allow usability and interoperability

In referring to 'high and common standards' the question addresses two complementary topics:

- Common standards: collecting and recording data in consistent ways.

- High standards: collecting data that is of a high quality – i.e. complete, valid, accurate and timely.

Taking these in turn:

1. Common standards

There are some areas where cross government standards could be established which would be of some benefit. Examples of this are:

- The way in which routine data sets are anonymised and pseudonymous identifiers attached to records (to allow linking within data sets and/or linking across data sets).
- The way in which survey data is released

In many cases, more important will be defining of standards for key data collections within specific areas. For example:

- The way in which primary care data is recorded by GP practices
- The way in which pupil progress is recorded by schools

These are issues where a competent authority is needed within each area.

2. High data standards:

High data standards should be driven by imposing duties on organisation recording data to record it accurately. Recording of data about public services is increasingly important in driving the way in which performance is measured and the way in which public services are paid. Consequently, the temptations to mis-record and the implications of mis-recording information are increased. The legal duties of public sector organisations and the responsibilities of directors and officers of such organisations to ensure information is accurate need to be strengthened.

These initiatives should be further supported by professional organisations in education, healthcare, policing, social care and other public services.

Q2.2 Is there a role for government to establish consistent standards for collecting user experience across public services:

Yes. There is scope for expanding the use of standard survey instruments across public services. However, the real benefit of this will only be achieved if the results are made more widely available than is currently the case.

Collection of user feedback in public services is in its infancy. It is perhaps most developed in healthcare where there are competing approaches - e.g. Patient Opinion and I Want Great Care.

It should be routine practice for public sector organisations to engage in this activity. However, different public services are at different stages in this process. In healthcare, where it is now common practice, an increased level of standardisation would bring benefits. In education, where such activity is non-existent or in its infancy, the emphasis should be on encouraging greater activity.

Q2.3 Should we consider a scheme for accreditation of information info-mediaries and how might this best work?

No system should be created to accredit the outputs for organisations using public sector data, in terms of the accuracy and reliability of information. This should be left to external agencies. Debate about the correct way to use data to better understand the performance of public services is fraught with conflicting viewpoints that are equally legitimate. The area is highly technical. Attempts by the government to “approve” certain uses of data and “disapprove” of others will require a vast effort that will satisfy very few and limit legitimate debate.

Clear standards to which organisations must comply in order to access anonymised data where a risk of re-identification remains are needed. A system of prior-accreditation could be introduced to establish organisations that meet such criteria. However, such a system should be kept to a minimum and must be one which enables organisations to be accredited promptly.

Q3. 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.

3.1 How would we ensure that public service providers in their day to day decision-making honour a commitment to open data, while respecting privacy and security considerations?

This can be achieved through the establishment of standards of service to external organisations seeking data. These standards of service should include maximum time frame within which data requests will be responded to; a transparent process by which applications for data can be made; a right of appeal in the event of applications being turned down.

These standards should be established first for national government departments holding key data sets and then expanded to local organisations. The key government areas are listed below under 3.4

3.2 What could personal responsibility at Board-level do to ensure the right to data is being met? Should the same person be responsible for ensuring that personal data is properly protected and that privacy issues are met?

Responsibility at board level should cover making data accessible, quality of data and ensuring privacy is protected. These functions can sit with one person as long as that person has clear rules around the issues of privacy and accessibility.

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

Yes

3.4 What sectors would benefit from having a dedicated Sector Transparency Board?

Health, Education, DWP, Home office.

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?

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

Data inventories should be organised in terms of value, starting with national government departments and national government agencies, then moving to local government and finally to local public service delivery organisations.

Data inventories should priorities giving details of:

- data sets that are continuously collected each year
- data sets that are mandatory nationally
- opinion survey data

Data inventories should identify:

- The data fields with definitions in each data set
- The organisations submitting data

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

See above

4.3 In what areas would you expect government to collect and publish data routinely?

See above

4.4 What data is collected ‘unnecessarily’? How should these datasets be identified? Should collection be stopped?

The main areas of waste are:

- a. Processing of data by central agencies.

This is the production of wide ranges of data tables for public consumption, the creation of websites to disseminate these tables, the planning of future releases of data tables etc. This activity is, in the main, of limited value since it is not driven by the needs of data users.

- b. Poorly designed data collections – e.g. the NRLS

There is considerable processing of data to report on government targets. This involves polishing of data and application of complex exception rules to ensure target measurement

is fair. The value of this activity depends on your view of the value of the target but is often limited.

4.5 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

Government should distinguish between "official statistics" and data releases. Polishing official statistics and ensuring they are of high quality is appropriate. However much of this activity currently is of little or no value. In contrast, data release should be carried out without first 'polishing' data and should not be limited by concerns about data quality. Data release should acknowledge that potential weaknesses in the data. However release will enable identification of data quality problems sooner.

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

5.1 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?

A central portal is preferable but either way, the thing of most value is cataloguing data and applying metadata in way that makes searching for and finding relevant data faster.

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

Priority should be given to those data sets that will meet the six objectives set out in the consultation. In particular we would recommend prioritising those that will support improvements in productivity of public services and economic growth.

5.3 What is more important: for government to prioritise publishing a broader set of data, or existing data at a more detailed level?

Existing data at a more detailed level. The key public service datasets need to be available in anonymised formats (see above).

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?

6.1 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?

Use of data by public sector organisations, professionals and informatics businesses does not require government stimulation. There is sufficient economic value in these activities that the market will respond.

Use of data in ways that engage the public, support social growth, accountability and choice would benefit from support. The media will provide some stimulus to choice and accountability. However,

UNCLASSIFIED

the use of data to promote social growth is an area that is likely to receive least attention. Government initiatives to stimulate such activity would be beneficial.

UNCLASSIFIED