TSO
information & publishing solutions

Making Open Data Real

A Public Consultation

October 2011

TSO

TSO is the leading provider of publishing services to the public sector and is at the forefront of working with public sector clients in open and linked data projects.
www.tso.co.uk
http://openup.tso.co.uk

Email response to

opendataconsultation@cabinet-office.gsi.gov.uk

Open data consultation

We welcome this consultation and see open data as an increasingly important part of the government's toolset for achieving increased efficiency and effectiveness for public services as well as boosting the growth of information re-use and re-publishing parts of the British economy.

There is, however, a clear distinction which can be drawn between data which will benefit citizens, government and industry by becoming openly available under a public right to data and information licensed by public sector organisations which has been developed under devolved activity, creating content around data which generates income to government. This type of process provides 'added value, message or meaning' and is often created on a commercial model. Overall we believe that this model saves government cost in its content creation and publishing and delivers a commercial return to government which would be at risk if it was made available free.

Where the costs of publishing or releasing of data are not judged to represent value for money, then the Public Data Corporation should have a role in finding ways to open up the data. Options may include the requestor paying for the data or paying for the transformation of the data to the required format. If the requestor pays then they could have preferential access to some or all of the data for a period of time. The transformed data could then be made available to other users either free of charge or on a freemium model. Other options such as partnership models and sponsorship should be considered.

To encourage publication of data by public service providers we need to encourage the data owners to see the benefits in opening up data – see Meaningful Open Data and Government Sets the Example, and provide standards and guidance – see Setting Transparency Standards.

An enhanced right to data

We support a stronger presumption in favour of publication than currently exists. Public service providers should be required to assess all their data for publication within clearly set guidelines. It should be part of the Public Data Corporation's responsibilities to ensure data is published, check that it is published to appropriate standards, assess the commercial value of datasets and help public service providers to find a way of opening up the data.

To ensure Open Data standards are embedded in new ICT contracts, government will need to set preferred open data standards within those new ICT contracts for different types of data. To ensure consistency across datasets, schemas could be created for different types of government data.

Setting transparency standards

We support a 5 star approach which aims for key government data to be published at 4 star or 5 star level in line with the public data principles. Publishing in this form will make data more accessible and enable re-use both by government organisations themselves, to improve service quality, and by third parties who can create more useful information sources from it (see meaningful data).

To help achieve this, standards for capturing the data so that it can be more easily transformed may need to be considered. Alternatively data can be 'harvested' from databases or files in an automated way and transformed into multiple formats.

Data published at 1, 2 and 3 star levels may not be particularly convenient for re-users, in which case further transformation work to deliver linked data will be required.

There is a role for government, perhaps through the Public Data Corporation, to set standards to help data owners identify which types of data should be opened up to which data standards. Data owners should be encouraged to aim for a 4 or 5 star approach, not just 3 star. When publishing data, the data owners should be able to explain the reason for the star rating of the data.

Under the Information Fair Trader Scheme any organization licensing information has to be independently verified. A similar approach could work for releasing public data to required standards.

An accreditation scheme for information intermediaries, or a framework, could enable the creation of an approved group of suppliers to provide services to both data owners to help them open up and commercialise data, and data re-users.

The benefits of an accreditation scheme for information intermediaries include:

- A clear procurement channel for data transformation
- Assurance of quality levels and proven capability in data transformation and resupply of data
- Sustainable data publishing processes and robust infrastructure, which will help data re-users to trust that data and encourage re-use
- Better understanding of user needs where the main users differ from originating departments

Services provided by information intermediaries could include:

- Creating schemas
- Transformation of data
- Resupply of data
- Hosting of data
- Re-using data

Corporate and social responsibility

We would suggest that part of the transparency process should be for organisations to publish a statement of how they have reviewed their data and what the rationale has been for opening or not opening that data. This information, along with a timetable for opening up data should be published on data.gov.uk or through the Public Data Corporation.

This approach would be preferable to a sanctions framework.

Meaningful open data

The best way to get public service providers to understand the value of open data is to get them to open it up and re-use it themselves. To help encourage organisations to publish datasets there needs to be a focus on the benefits to government in being able to use that data themselves to improve services – it is likely that data will be useful to external audiences as well.

For example, many organisations have data in silos which cannot be related easily. Using semantic techniques to link those datasets may help to drive more value from existing datasets which could help to inform how to deliver better services. A good example was the Govspark initiative which tracked energy usage across central government buildings, enabling high usage to be recognized and actioned to bring down cost and carbon emissions.

Another example of where using semantic technology has helped to reduce cost is the expert participation model for legislation where transforming legislation into machine readable linked data has enabled automated updating processes which will reduce the cost to government in the long term.

If publishing data in a machine readable, re-usable way became the norm then government could use its data more effectively to reduce costs and improve efficiency. For example, FOI requests could either become easier to answer or eliminated if the data has been published externally already.

Prioritization of transformation of datasets into linked data formats (i.e. 4 and 5 star data formats) could be around those datasets which will be most useful for public sector organisations, thus funding the transformation work by the longer term cost savings. Other datasets should be prioritized based on external demand.

Although good progress has been made in publishing key datasets, publishing them is just the first step. In several cases, complex datasets need more work to make them fully usable and useful to a wider audience. This can be achieved by commissioning third parties to do the necessary transformation, either paid for by government or by third

parties that will benefit from the transformed data. Creating sustainable data publishing processes that capture data in a way which makes it easier to be transformed automatically will help to create a forward plan for opening up more datasets and delivering reliable updates to the data.

Data should be updated regularly and the frequency of those updates should correlate with the use of that data. To enable this, the process needs to be automated as much as possible. Sustainable data publishing processes need to be put in place which will streamline the capture, transformation and publishing process, delivering accurate, up-to-date, re-usable data on robust infrastructure.

Data quality should be appropriate to the data and use. In defining what is high quality data, the required frequency of updates and the ease of re-use need to be defined.

Government sets the example

As stated above, the best way to get government to understand the value of open data is to get them to open it up and reuse themselves. From experience, we know that in many cases the largest use of government information is government and being able to access and combine public service data in a consistent way will help to improve efficiency. Demonstrating that there are benefits in publishing data to a high standard, both in terms of unlocking value from existing data silos and encouraging useful data applications to be created by third parties, will encourage government organisations to release datasets. Efficiency savings and revenue should be measured and reported to encourage further transparency.

Innovation with open data

A government publishing model already exists which helps government to productise government guidance and drive innovation and value from that information at zero cost to government. By licensing key datasets to commercial companies the same model could be applied.

There may also be a role for government to stimulate enterprise by selectively opening up strategically important data, i.e. where there is significant value to third parties and the general public, to 5 star standards. For example, developing the legislation.gov.uk website based on open data principles has enabled the creation of third party applications as well as delivering efficiencies and ultimately has created a best practice example for publishing government information.