

Introduction

The British Library welcomes the opportunity to submit its views to the Cabinet Office consultation 'Making Open Data Real: A Public Consultation'. The consultation paper sets out Government's proposed approach for a Transparency and Open Data Strategy, which is aimed at establishing a culture of openness and transparency in public services.

Background to the British Library

The British Library was established by statute in 1972 as the national library of the United Kingdom. It is one of the world's greatest research libraries - it benefits from legal deposit and is the main custodian of the nation's written cultural heritage. The Library's incomparable collections, which contain more than 150 million items, have developed over 250 years; they cover three millennia of recorded knowledge, represent every known written language, every aspect of human thought and a considerable sound, music and recordings archive.

The Library occupies a prominent place in the intellectual and cultural life of the nation. It is an integral component of the research infrastructure and it plays a correspondingly significant role in ensuring the research excellence of the UK. As a catalyst for creativity and innovation it supports economic growth and we aim to be at the heart of the knowledge-based economic recovery. Sir Isaac Newton said: "If I have seen further it is by standing on the shoulders of giants". This is what the BL seeks to assist its users to do. The Library's mission is to advance the world's knowledge. The Library serves five principal user groups – researchers, the business community, the UK library and information network, learners and the general public. In 2010/11, more than 10.3 million British Library collection items were consulted by, or loaned to, academic researchers, business researchers, and private individuals. An independent economic impact study commissioned by the British Library suggests that the total value added to the UK economy by the Library each year is £363m, or £4.40 for every £1 of public funding.¹ Up to 6.5 million people visit the British Library website - www.bl.uk - every year. The Library also presents a wide range of free exhibitions, public events and publications, to broaden access to, and enjoyment of, its collections, attracting approximately 100,000 visitors each month. In addition the Library's Learning team proactively engages with schools and lifelong learners, offering workshops for 20,000 visitors and 1.2 million online users each year.

¹ *Measuring our Value: Results of an independent economic impact study commissioned by the British Library to measure the Library's direct and indirect value to the UK economy* (December 2003).

The British Library and Open Data

The fundamental *raison d'être* of the British Library is precisely to facilitate access to the world's knowledge base, particularly in the digital environment, and to lead and collaborate in growing the world's knowledge base. It is in that context that we would draw a distinction between the data created precisely to fulfil our core purpose of supporting research, creativity, innovation and economic growth on the one hand; and on the other, data collected as a by-product of delivery where the openness of that data will, as the consultation paper underlines, principally support accountability, transparency, and other applications. It is with the open data in the first category that we are most engaged and where we seek to take a distinctive and leading role.

With regards the former category, the British Library's open metadata strategy is designed to remove barriers and enable increased innovation without imposing unnecessary restrictions. This resulted in the British Library announcing in August 2010 that it would be opening up its rich set of bibliographic metadata for re-use.

Since then, the Library has signed up over 400 organisations in 69 countries to a free catalogue data service; created a linked open data version of the British National Bibliography; become one of the first signatories of the [JISC Discovery Open Metadata Principles](#) and started to offer sets of metadata to researchers under a permissive Creative Commons license.

The Library's metadata strategy involves millions of its catalogue records being released using the Resource Description Framework (RDF) format for use by the wider community under a highly permissive Creative Commons Zero (CC0) license (See: <http://www.bl.uk/bibliographic/datafree.html>).

The Library's free data services have included the creation of a linked open data version of the British National Bibliography which consists of nearly 3 million items (bnb.data.bl.uk/sparql) and meets the 5 star rating noted in section 8.9 of the consultation document. This service is currently processing over 850K user transactions per month.

In addition to the Library being a signatory of the JISC Discovery Open Metadata Principles, the Library is working with other partners including Europeana and the Conference of European National Librarians (CENL) on a number of initiatives involving the opening up of library data for wider use. The Library is therefore an organisation that has both experience of, and a keen interest in the technical, licensing and strategic issues involved in the widening of access to public sector data and metadata. In addition the Library also has an interest in the implications of any proposals for public sector organisations with data or metadata that is not directly covered by central or local government regulations (e.g. in the areas of Crown Copyright or the Open Government License).

These developments have been widely applauded in the international community, and the Library has taken a leading role in Europe on this issue. As a result, in September 2011, the Library welcomed the announcement made by the Conference of European National Librarians to support the open licensing of their data. This will mean that datasets describing millions of books and texts published in Europe will become increasingly accessible for anyone to re-use for whatever purpose their wish.

With regards the latter category, as a public body we comply with the Government transparency and accountability agenda and [the British Library](#) already contributes several data sets to data.gov.uk including:

- [Spend over £25,000 in the British Library](#)
- [Contract spend over £10,000 by the British Library](#)
- [Organogram and staff pay data for the British Library](#).

The British Library's responses to the individual Consultation Questions follow overleaf.

The British Library
October 2011

Responses to the Consultation Questions

Introduction/Glossary (p.6)

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

Dataset: While the definition seems clear, the narrow scope is somewhat at odds with the wider ambition of the 'vision'. Is this definition intended to be limited to 'management information' or does it include other data or metadata produced by public bodies? For example is it also intended to include: datasets produced by publicly funded research; Ordnance Survey data, bibliographic data or user contributed (Web 2.0) data? If so, some rewording or clarification may be required.

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

In principle the initial test must be that of legality (i.e. third party licensing, IPR or privacy issues do not apply to the data under consideration for release). However, in practice the decision also has to take account of the technical, financial and support issues involved. The balance to be struck is one between public interest in the data and the various resources required to make it available in the format required e.g. data may need to be digitised or converted.

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

Public sector organisations could reasonably be expected to support a clearly defined and publicised set of base level data offerings compliant with both sector and wider interoperability standards. Payment would become an option if a requestor required significant additional work on the base level offering to support creation of a unique customised data file e.g. filtering, reformatting, merging etc prior to supply.

4. How do we get the right balance in relation to the range of organisations (providers of public services) our policy proposals apply to?

A combination of basic cross-sector guidance together with targeted requirements for specific groups (e.g. use of sector standards) should assist in the creation of comprehensive but appropriately weighted coverage. In order to minimise confusion or wasted resources and maximise take up, clear requirement definitions concerning the applicability of measures to different types of public sector organisation will be required e.g.:

- Central government
- Local government
- Non-Departmental Public Bodies etc

Groupings (e.g. cultural, financial or medical).within these categories which are of relevance to those aggregating data should also be clearly identified in order to create well balanced organisational coverage together with provision of the most relevant and useful datasets.

What threshold would be appropriate to determine the range of public services in scope and what key criteria should inform this?

One basic criterion might be an established threshold level of public funding to the service to be set against the independently established public interest or utility of the data involved.

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

There are recurring themes in data management regarding responsibility, accountability and incentives for sharing. Creation of open data can be a complex and demanding process for organisations without a history of sharing electronic information with third parties, particularly those beyond their organisational sector. There is recognition that specialised and embedded skills are required for the efficient and effective management of data. Presently, approaches to providing the necessary training and infrastructure to support this need are patchy. Many organisations additionally consider the financial burden for data maintenance and sharing to be an issue. This situation is exacerbated by a lack of costing models for data conversion and validation activities. There is therefore scope for government to coordinate development of robust and comprehensive data sharing models and best practice guidance for use by the public sector.

In order to promote wider data publication effort will be needed to ensure all requirements and support processes are expressed as unambiguously as possible with:

- Clear sectoral guidance provided on:
 - Standards
 - Licensing
 - Required service levels, KPIs, benchmarks etc
- A centrally maintained data repository for organisations to offer data supporting standard access technologies and formats
- Defined guidance or best practice on required recognition or attribution of contributions
- Use of all appropriate communication channels i.e. web site use cases, expert workshops, developer hack days etc to assist:
 - Dissemination of public sector experience best practice of data release
 - Improved contacts between releasing organisations and developer or entrepreneurial communities
- Technical support resources covering:
 - IPR, licensing issues and liability etc
 - Data formats, conversion issues, etc
- Identification and promotion of benefits for organisations supporting their data release e.g.
 - Improved visibility
 - Increased perception of public value e.g. from attribution or citation of organisational data
 - New technical collaboration opportunities etc.

1) An Enhanced Right to Data (p.25)

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

Public policy formation in the area of open data needs to strike a balance between protecting public and individual interest, maximizing opportunities for knowledge creation or sharing and inspiring entrepreneurial activity for economic benefit. The creation of a stronger presumption for publication of data will require a combination of measures including:

- Formal requirement for public organisations to offer data (e.g. similar to FOI?)
- Communications to data consumers and the wider public on the available opportunities for using public data including a centralised point of enquiry for those searching for data
- Lower barriers for organisational data release via support to public sector organisations (e.g. as listed in response to Q5 above) enabling a greater choice of available data.

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?

The enhancement of the existing role of Information Commissioner to incorporate wider powers covering open data would involve significant additional responsibilities and new areas of expertise. A new specialist unit within the Information Commissioner's Office may therefore be a more appropriate option. However, considerable care will be required to ensure that the ICO is adequately resourced and prepared to support what would inevitably be a complex start up process for both the ICO and public sector bodies.

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

No. Identification of links between previously disparate data elements in order to create new information or applications is one of the key benefits of linked open data. Cross correlation of data from multiple organisations therefore offers new opportunities to infringe privacy in ways that individual suppliers would be unaware of since:

- Their data was not originally designed with such usage in mind
- They may be unaware of how other data offerings might be linked to their data
- Possibilities for linking may develop over time and thus data that originally offered limited scope for correlation has richer opportunities for connection as further sources or tools emerge.

Considerable thought and technical advice will therefore be required to ensure:

- Adequate protection of individuals or organisations
- Accurate advice to public sector organisations on how to prevent unintended consequences of published data e.g. via anonymisation
- 'Take down' or other remedial options to negate organisational liability in disputed cases.

4. What might the resource implications of an enhanced right to data be for those bodies within its scope?

Organisations responsible for the supply of open data would inevitably face a number of new and potentially complex resource challenges including:

- A requirement to adopt and utilise new standards and technologies to enable wider usage and interrogation of their data
- Support of enquiries and technical queries relating to data releases
- Creation of suitable open platforms or channels for delivery of data - if no central government or sectoral repository were to be provided
- An ongoing requirement to upgrade formats and technologies in line with external drivers rather than inhouse requirements (i.e. developers may expect compliance with current or cutting edge systems or standards when institutions may not yet be able to support these due to resource constraints).

How do we ensure that any additional burden is proportionate to this aim?

If open data is to be a real driver for economic growth and a free market approach promoted, a considerable degree of standardisation will be required in order to maximise opportunities for re-use. As previously noted, lowering data publication barriers via the promotion of well publicised standards, free open tools and a centralised service platform would help to lower the additional overhead for data production. Attention should also be paid to the management of user expectations e.g. through the clear promotion of common standard data offerings with supplementary advice on priced added value options when available.

Consideration should also be given to establishing an upper cost limit similar to FOI for data requests (*see Q3 Section A above*) in order that prohibitively expensive examples may legitimately be rejected where appropriate. However it should be noted that data extraction or export may be considerably more complex than those of information gathering and thus longer timescales and higher limits may be required to reflect this.

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

The issue of incorporating open standards into ICT procurement exercises is a complex one requiring appropriate weighting of the objective to share data against many others involving the fitness for purpose and best value of the application solution. However, a number of measures might be undertaken to actively promote the inclusion of open data as a tender requirement including:

- The capability of systems to manage and export data to open standards should be required in ICT procurement scoring models.
- The government agency with responsibility for promoting the open data agenda could:
 - Work proactively with suppliers to improve awareness of public sector open data requirements and related standards and emphasising that features supporting these objectives would be scored more highly in procurement exercises

- Provide public sector organisations with advice on systems procurement best practice involving open data issues e.g. avoidance of systems or third party data enhancements that may have proprietary features
- Provide additional advice specifically on the incorporation via licensing of added value third party data into core institutional data sets with the aim of not compromising the viability of wider or long term open data release due to short term IPR confusion.

2) Setting Open Data Standards (p.28)

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

While no single factor can easily be isolated to ensure standards compliance, the cumulative effect of a number of coordinated factors could assist their adoption and usage including:

- Promotion of non-sectoral, wider standards offering established practice for data suppliers and reduced barriers for developers and users.
- Provision of information on suitable tools and infrastructure to support data conversion
- Encouragement of the use of public sector controlled vocabularies to facilitate interoperability and consistency.
- Guidance on a standard approach to licensing & usage terms
- Consideration and weighting of open data issues in procurement exercises.

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

While this approach may be effective when evaluating end user systems or services, collection of usage information for open data can be more problematic since open licensing may mean data is passed on to third parties by the original requester. Collection of accurate user information may therefore be limited only to first generation users. In addition data may be blended with further information to generate the final result. Possible measures government might employ to assist in this area could include:

- Suggested best practice on enabling collection of user feedback
- Preferred models for demonstration of public value of data release
- Centralised mechanisms for collecting feedback on user experience possibly linked to a public sector data sharing platform.

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

While there may be some value in identifying information intermediaries with proven track records, the rapidly changing information landscape may ensure such accreditation dates incurs a significant burden of administration. In addition, if one aim of the government's open data release initiative is to encourage innovation by UK start-up companies with good ideas but no previous track record then an over-emphasis on the use of accredited companies could compromise this objective.

3) Corporate & Personal Responsibility (p.30)

- 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.**

If the intention is to embed an open data culture in the public sector, then the commitment must become a business as usual responsibility similar to that relating to Freedom of Information and Data Protection with clear and unambiguous instructions relating to organisational responsibilities.

In the area of privacy and security, the British Library's anecdotal evidence is that individuals are aware of privacy and data security issues. We believe that clear lines of rights, responsibilities and accountability, particularly for individuals who work with sensitive data, provide a solution to most issues surrounding privacy and security.

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

Introducing a personal responsibility at Board-level would undoubtedly underline the relative priority of the government's open data agenda; however in the Library's view very careful consideration indeed would be needed to ensure both proportionality and also consistency with the allocation of the range of other corporate responsibilities at Board and Senior Executive level.

- 3. Would we need to have a sanctions framework to enforce a right to data?**

Whilst a robust sanctions framework would also undoubtedly underline the government's strong commitment to open data principles and ensure appropriate prioritisation in strategic planning, the Library would be concerned that embedding initiatives be balanced with proportionality and that application of any sanctions should be a last resort option and should be equally balanced by measures to positively encourage public sector action.

- 4. What other sectors would benefit from having a dedicated Sector Transparency Board?**

The Library does not have a view on this issue.

4) Meaningful Open Data (p31)

- 1. How should public services make use of data inventories?**

Institutions should use data inventories as a centralised record of all data and metadata collected, acquired or generated by the organisation together with details of:

- Technical details (e.g. format & standards used)
- Availability status (e.g. currently available, available with further technical work or unavailable due to licensing or privacy issues etc)
- Context & supporting information.

What is the optimal way to develop and operate this?

Data inventories should be maintained by individual organisations responsible for the data in a form compatible with export or upload to any centralised listing (e.g. data.gov.uk) used for public access. Technical guidance on the structure and format of such inventories will therefore be required in order to develop and maintain compatibility and prevent the collection of unnecessary data.

2. How should data be prioritised for inclusion in an inventory?

A number of factors might be used to assist the prioritisation process including:

- Public interest
- Research value
- Relevance to specific communities or collaborative projects (national, international or specialist)
- Potential for commercial exploitation.

Such factors will need to be weighted appropriately against potential issues such as:

- Overall development effort required
- Technical complexity e.g. format/standards issues
- Additional requirement to filter data to protect privacy issues
- Support and frequency of updates e.g. to maintain relevance / currency.

How is value to be established?

A combination of measures will be required relating to:

- Verification of the prioritisation factors noted above (i.e. relevance/interest to researcher, commercial exploitation etc)
- Feedback from users
- User survey of those utilising the data
- Usage figures from linked open data services
- Numbers of systems, applications utilising the data
- Economic benefit generated for the UK economy.

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

Areas of data collection will inevitably vary within individual sectors however; common organisational data might be expected to include:

- Performance against area / sector KPIs (e.g. environmental, health & safety etc)
- Financial data on topics covered by the organisation's annual report
- Organisational statistics
- Supporting data for special projects, inquiries or other initiatives.

4. What data is collected unnecessarily? How should these datasets be identified? Should collection be stopped?

The value of data collection should be subject to similar checks as other activities involving the use of public resources i.e. there should be an established local business case for data collection underpinning the efficiency of the collecting organisation in undertaking core functions or responsibilities. Resources should not be wasted on collecting data for which no such case exists.

5. Should the data that government releases always be of high quality?

There is a compelling case for establishing a hierarchical series of data definitions similar to the 5 star model suggested in the consultation that can be used to indicate the potential compatibility and utility of the released data. However, quality should not be used as an excuse for withholding data. If there are known issues with the format, accuracy or currency of the data these should be made explicit to potential users. If for example data complies with local rather than wider standards this should not automatically prevent publication; instead such examples should be accompanied by documentation covering the local standards used until more interoperable data becomes available.

How do we define quality?

A useful definition of quality is fitness for purpose. However, in the case of open data the user's purpose may well be unknown to the data creator. There are however some common key features that are likely to be valuable including:

- Compliance with documented open or international standards e.g. Unicode
- Consistency of content
- Minimal spelling or other basic errors
- Provenance information (critical in establishing trust)
- Data definition or scope including:
 - Date of collection
 - Relation to other data collected
 - Geographical coverage
 - Subject area
- Established purpose of collection or creation
- Adequate availability of data format details and supporting information
- Support information for further queries.

To what extent should public service providers polish the data they publish, if at all?

While the meaning of core data content should not be modified, consideration should be given to:

- Clarity of data labels for non-specialist audiences
- Suitability of data for cross sector comparison
- Interoperability of data content and format
- Availability and suitability of supporting documentation for non-specialist audiences.

5) Government Sets the Example (p.33)

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?

Availability of public open data from a single centralised source would assist:

- Researchers - via presentation of a 'one stop shop' standard service offering
- Data creators - by simplifying processes and lowering barriers to data re-use
- The government open data agenda - by minimising duplicated effort, improving consistency of presentation and improving the speed of take up.

Within a centralised repository data could still be presented grouped by department or sector e.g. culture, sports etc, according to user preference if the system were suitably configured.

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

In general terms prioritisation should be linked to:

- Wider public and research interest beyond organisational activity (e.g. reusable location, weather or accident data)
- Potential for data to be exploited to generate UK economic growth
- Relevance to wider communities e.g. utility for linking with other data to create new hybrid services (e.g. addition of mapping data can multiply the utility of many other datasets when linked)
- Lack of issues relating to IPR, security, privacy requiring some additional work
- Low development costs to enable export of data in a reusable form.

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

For the foreseeable future continuous iterative improvement will be required to make most public data fully interoperable. Prioritisation may therefore require a case by case analysis dependent upon the relevance and utility of the data involved to enable useful outcomes. Cross comparison of sectoral data necessitates the availability of a representative data selection to compare. Thus, preference in this instance might for broader coverage. In other cases, specific data sets may be of unique interest and thus prioritisation over others may be appropriate. In both cases however data would still need to be suitably detailed to enable useful comparison or analysis.

6) Innovation with Open Data (p35)

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?

In a global environment it is increasingly likely that opportunities for innovation may not be solely exploited by UK based companies. Careful thought should therefore be given to if and how UK entrepreneurial activity might be best supported in the face of wider interest?

There are a number of proactive measures that can be used to stimulate awareness of and experimentation with open data e.g.

- Resourcing of:
 - Support web sites
 - The creation of suitable data manipulation tools and related documentation
 - Awards for best implementation of public metadata
 - Data hack events with software developers etc
 - Targeted conference events & communications
- Creation of opportunities for the use of data in projects or education activities in UK schools and universities.

It is likely for this type of activity to be fully successful resources will also be required from organisational data suppliers in order to provide more detailed support to users/developers etc. This additional requirement should therefore be factored into any institutional cost equation.

The British Library
October 2011