

<b>Title</b>	Response to 'Making Open Data Real' consultation.
<b>Revision</b>	2
<b>Date</b>	24 <sup>th</sup> October 2011
<b>Author</b>	Paul Davidson, LeGSB Director of Standards, and CIO Sedgemoor District Council
<b>Purpose</b>	LeGSB response to consultation.
<b>Subject (IPSV)</b>	1526-Information management
<b>See also</b>	<a href="http://www.cabinetoffice.gov.uk/resource-library/making-open-data-real-public-consultation">http://www.cabinetoffice.gov.uk/resource-library/making-open-data-real-public-consultation</a>

Revision	Date	Author	Notes
1	24/10/2011	Paul Davidson, for LeGSB	
2	27/10/2011	Paul Davidson	Reference to the separate 'Linked Data Experts' submission.

## This document

This document provides responses to the questions raised in the 'Making Open Data Real' document at

<http://www.cabinetoffice.gov.uk/resource-library/making-open-data-real-public-consultation>

Given our role in eStandards, we have focussed on the questions where standards can play a role to achieve the best re-use of open data. We hope that our input will make a positive contribution to making open public data real and useful.

We have also been happy to contribute to a separate collective response from 'Linked Data' experts.

## About the Local e-Government Standards Body (LeGSB)

A description of LeGSB and its constitution can be found at [www.legsb.gov.uk](http://www.legsb.gov.uk). In summary

- LeGSB was originally an e-Government National Project. Since 2006, the accountable body is Tameside Council.
- LeGSB's mission is to promote eStandards that support Efficiency, Transformation, and Transparency of Local Services
- LeGSB is currently funded from the Cabinet Office, Department for Communities and Local Government, and the DWP.
- LeGSB's Chair is John Curtis from Mersey Fire and Rescue and the Director of Standards is Paul Davidson, CIO at Sedgemoor District Council. The LeGSB Board has representation from Local Authority ICT managers, DCLG, DWP, Cabinet Office, LGA, SOCITM, Health, Education, and Intellect.
- Paul Davidson represents English Local Authorities to the Cabinet Office CTO Council and its Information Domain where public sector eStandards are devised within the ICT Strategy and the Standards and Architecture Framework.

## Responses to the Questions posed in the Consultation

	LeGSB Response
What is the best way to achieve compliance on high and common standards to allow usability and interoperability?	<p>Compliance will be a combination of Incentives and mandation.</p> <p>Incentives</p> <ul style="list-style-type: none"> <li>• Make standards attractive to use so that they bring benefits to the organisation that is implementing them. Those benefits could include</li> </ul>

- Efficiency Gains
- Greater Insight
- Improved Data Quality
- Access to new data sources
- Make standards 'findable' for a scenario. A single point to find the right standard for a scenario/subject/purpose.
- Make standards easy to implement. This could include
  - Tools
  - Knowledge Base
  - Translation services
  - Support
  - Skills
  - Collaboration

#### Mandation

- Dictate the standard to use when publishing data for a new given subject/purpose.
- Where a standard does not exist; expectation for a representative group to collaborate to generate a standard that will be universally adopted for a given subject/scenario.

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

The Open Public Services white paper pre-supposes that a range of data and information services are available ...

...being totally transparent about the quality and value for money of public services so that new providers can come in and challenge under-performance.

People will be able to choose what sort of service they want and find the best provider to meet their needs.

... that people have access to the information they need to make good choices and hold providers to account.

... those choosing a service are well informed and prompted about the options available;

... we will ensure that key data about public services, user satisfaction and the performance of all providers from all sectors is in the public domain in an accessible form. This will include data on user satisfaction, spending, performance and equality

**Open Public Services white paper - 2011**

This type of distributed information about services, choices, performance, and satisfaction can only be achieved when the data is published to a standard so that it is query-able and comparable. While some of these standards could be left to the private sector to develop, that may lead to competing standards and regional solutions. The reference data that drives this ecosystem originates within the public sector, for example, what organisations are responsible for provision of what services covering what area.

So – government should invite proposals for standards that cover this scope and provide the leadership to consistently implement a cohesive set that come together to achieve these outcomes.

The Government ICT Strategy ( March 2011 ) commits to implementing an Open Standards engagement process (led by DoH) in which proposals for standards that should be applied across government can be made and assessed.

That strategy also commits to building a 'Reference Architecture' (led by HMRC)

	<p>which includes a data definition layer. These standards should link back to the 'Reference Architecture' so that they are discoverable and re-useable.</p>
<p>Should we consider a scheme for accreditation of information intermediaries, and if so how might that best work?</p>	<p>Not clear what is meant by an 'Information Intermediary'.</p> <p>If it is about those aggregating and combining public sector data, then accreditation does not seem necessary so long as the licence terms of each data set are complied with.</p> <p>If public sector services themselves ( including web sites ) are relying on these 3<sup>rd</sup> party aggregators, then they will need to be assured that the services will be sustained and operating to service levels.</p>
<p>How should public services make use of data inventories? What is the optimal way to develop and operate this?</p>	<p>Discovering public data on a subject, and having confidence that you have found it all, is going to be important to mainstreaming access to data.</p> <p>Public Sector Bodies should publish their data holdings ( not the actual data ) in both human-readable and machine-readable forms with consistent metadata to include</p> <ul style="list-style-type: none"> <li>• Subject</li> <li>• Coverage</li> <li>• Time</li> <li>• Terms by which they can be shared</li> </ul> <p>These can be distributed rather than centralised if standards are applied.</p>
<p>How should data be prioritised for inclusion in an inventory? How is value to be established?</p>	<p>A priority should be the reference data that can be used to join up other transactional data.</p> <p>Where a public sector body or association is the authoritative source of definitions or identifiers for a subject, they should be expected to publish that as open data.</p> <p>This will include identifiers for</p> <ul style="list-style-type: none"> <li>• Public Sector Bodies, Services, Schools, Expenditure Types, etc</li> </ul> <p>This data can then be used to link otherwise disparate public data about the same 'thing' or topic, and also invite people and businesses to create their data using the same identifiers.</p> <p>Some of this core reference data is not 'open', in particular addresses. The consultation on the Public Data Corporation proposes licensing options for this type of data.</p> <p>Data such as the National Address Gazetteer should be published as open data such that</p> <ul style="list-style-type: none"> <li>• The identifiers of each address are published</li> <li>• Sufficient metadata is associated such as the address and the post code.</li> <li>• Sufficient added value data is not published as open data so that it can be separately licensed to support a commercial model.</li> </ul>
<p>In what areas would you expect government to collect and publish data routinely?</p>	<p>Information about public services to support the ambitions of the Open Public Services white paper ( see above).</p> <p>This kind of 'rolling' data should be published via an API so that real-time and near-real-time information services can easily access up-to-date data.</p> <p>These data services should be published to service levels giving confidence</p>

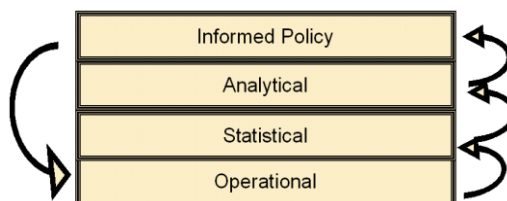
	<p>about the quality and longevity of the service. It is unlikely that businesses will build services on data that cannot be relied on.</p>								
<p>What data is collected “unnecessarily”? How should these datasets be identified? Should collection be stopped?</p>	<p>There are many examples of data that is collected by a part of government, that is already held definitively by a more appropriate part of government. This occurs when</p> <ul style="list-style-type: none"> <li>• The definitive data is not published</li> <li>• The definitive data is not promoted</li> <li>• The potential re-user does not trust the quality or longevity of the definitive data.</li> </ul> <p>This duplication should be addressed by expecting those that have definitive reference data and definitions to publish them as data and to promote them for re-use within government and beyond.</p>								
<p>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?</p>	<p>Some bodies are reluctant to publish data because it is not entirely accurate or complete or up-to-date and so on.</p> <p>There should be a consistent way of describing the quality characteristics of all data sets, using common terms and vocabulary.</p> <p>Quality Characteristics appear to be different for each context of data, for instance</p> <table border="1"> <thead> <tr> <th>Context</th><th>Example Quality Characteristics</th></tr> </thead> <tbody> <tr> <td><b>Operational Data – e.g. Cases</b></td><td>Measurement Tolerance Verification applied Authentication strength used</td></tr> <tr> <td><b>Statistical Data – e.g. aggregated data</b></td><td>Completeness Timeliness Accuracy</td></tr> <tr> <td><b>Analytics – e.g. conclusions from statistics</b></td><td>Strength of Association Specificity Plausibility Coherence</td></tr> </tbody> </table> <p>Quality Characteristics can become easily disconnected from the data set that they relate to. If data were published as ‘Linked Data’, the associated quality characteristics remain connected to each ‘cell’ even when it is downloaded.</p> <p>The expectation is that when quality can be communicated alongside data, organisations will be more willing to make it available as open data, and consumers will be more informed about the valid uses it can be put to.</p> <p>Data is likely to be fit for the purpose for which it was collected. For instance, a person may be asked for their data of birth as they apply for a library card. The library may choose not to check this. That library card should not then be used to prove age at a bar!. Data may be aggregated and used to profile library usage by age. The fact that the date of birth was not verified would probably be lost when re-purposed like that.</p>	Context	Example Quality Characteristics	<b>Operational Data – e.g. Cases</b>	Measurement Tolerance Verification applied Authentication strength used	<b>Statistical Data – e.g. aggregated data</b>	Completeness Timeliness Accuracy	<b>Analytics – e.g. conclusions from statistics</b>	Strength of Association Specificity Plausibility Coherence
Context	Example Quality Characteristics								
<b>Operational Data – e.g. Cases</b>	Measurement Tolerance Verification applied Authentication strength used								
<b>Statistical Data – e.g. aggregated data</b>	Completeness Timeliness Accuracy								
<b>Analytics – e.g. conclusions from statistics</b>	Strength of Association Specificity Plausibility Coherence								
<p>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?</p>	<p>LeGSB is working with Devon County Council towards publishing ‘linked’ information about their</p> <ul style="list-style-type: none"> <li>• Objectives</li> <li>• Strategies</li> <li>• Plans</li> <li>• Services</li> <li>• Partners</li> <li>• Resources</li> <li>• Stakeholders</li> </ul>								

	<ul style="list-style-type: none"> <li>Metrics</li> </ul> <p>etc, as Linked Data so that a, for instance,</p> <ul style="list-style-type: none"> <li>A commuter parent living in a given town in Devon can discover the strategies, resources, services etc that relate to them, rather than having to find and read through lots of documents, most of which are not relevant.</li> </ul> <p>This type of data should be published to a consistent standard, and using a consistent API, by each public sector body, so that it can be joined up by 3<sup>rd</sup> party dash-boards.</p> <p>The question mentions 'portals' which suggests that this type of information should be presented in that way. It should first be published as linkable data from which many portals representing a range of communities and stakeholders can provide a combined view.</p>
What factors should inform prioritisation of datasets for publication, at national, local or sector level?	<p>To date, the driving factor to prioritise open data has been scrutiny of public services and public life i.e. Accountability via Transparency.</p> <p>We should now have a focus on data for</p> <ul style="list-style-type: none"> <li>Participation in Local Democracy</li> <li>Supporting Innovation</li> <li>Reducing duplication</li> </ul>
Which is more important: for government to prioritise publishing a broader set of data, or existing data at a more detailed level?	<p>Selected existing data should be published as 'Linked Data'.</p> <p>As data is published with more detail, the need to define its semantics becomes more important.</p>
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?	<p>Promoting government reference data for external use would</p> <ul style="list-style-type: none"> <li>enable people and business to feedback to government about public services.</li> <li>'Crowdsourced' new data using government supplied reference data to build in links and comparability</li> </ul> <p>Government should be proactive in helping others to use public data</p> <ul style="list-style-type: none"> <li>Forming communities around certain datasets</li> <li>A collaborative environment to discuss a dataset and meet interested people</li> <li>A Knowledge Base for each dataset</li> <li>Mixing Public Sector Data with 'crowdsourced' data about the same topic.</li> <li>Giving the confidence to build a business around a dataset</li> </ul>

## General Observations

LeGSB has been pleased to work with the CTO Council's Information Domain to propose contexts for public sector information. We find these helpful to describe the applicability of some types of standards and information management techniques.

- Operational** – about real people and places, with real circumstances, needing real services, i.e. Case Work
- Statistical** – aggregated operational information, organized using common classifications.
- Analytical** – the conclusions drawn from the analysis of statistics. i.e. patterns, predictions, inferences, opinions.
- Political** – the decisions taken to shape services.



Applying standards that reach across these contexts can encourage an information sharing environment in which:

*The **decisions** that we take, are based on the **conclusions** that we reach, from the **statistics** that we gather, from the **data** that we handle.*

The Information Domain has used these contexts as a framework working towards the ICT Strategy deliverables for

- Information Strategy Principles
- Open Standards Engagement

Paul Davidson  
For LeGSB  
24<sup>th</sup> October 2011