

MAKING OPEN DATA REAL: A PUBLIC CONSULTATION

RESPONSE

Dr Robert Daniels-Dwyer, Enterprise Information Architect

27 September 2011

This response is based on my professional experience of working with major organisations in the UK on the development of joined-up data and information architectures, and in particular on my work for Network Rail and the Office of Rail Regulation between 2005 and 2009.

Overall, I find the government's commitment to 'open data' a very exciting and promising one, and would be happy to provide further views if they are of value.

Enhanced Right to Data: Q1

This presumption would only in practice be realised if there was a presumption that data in central and local government is stored in proper database management systems, or as a minimum in a centrally managed set of spreadsheets for each department. While tactical data store are produced by individual teams, the presumption of publication will simply fail. This is a greater change management issue than simply exhortations; there must be changes to processes and technology that direct people to share automatically unless they opt out.

Enhanced Right to Data: Q5

Most of the data captured by government is outside the scope of ICT contracts. There needs to be a government standard for data mark-up, such as xml, that is required for all contracts. The application of those standards can only be enforced by the maintenance of skilled teams within government. Approaches by the US Federal Government should be investigated.

Setting Open Data Standards: Q1

This is a change management issue. Departments data should be published to an agreed timetable, with data scored for example using the Berners-Lee scheme suggested. It becomes in the interests of departments to be seen to be publishing quality data, because it will be possible for the public (including the Opposition and the press) to produce league tables of the quality of publication by department.

Corporate and Personal Responsibility: Q1

There are sound constitutional and economic reasons for making public data available in the manner proposed in the Consultation Paper. The challenges that this approach will present to central and local government cannot be understated, however. There are obvious change management issues, which are addressed in the Paper, that public servants tend to see the data that they collect as being an internal matter. There are also significant technical challenges. Data within Whitehall and local government is frequently stored in *ad hoc* formats, such as Excel spreadsheets and Access databases, that have been designed for a specific use (which may be a one-off use) rather than for long-term and widespread sharing of data.

The Paper is well thought through. If there is a criticism, however, it is that it addresses the problem of making open data real as being a problem about taking data held internally, and publishing it to the outside world. I would suggest that the challenge is actually more fundamental than that: the civil service needs to be able to take the data that it holds in individual functional siloes, and make that data shareable, through the use of common (probably 'open') standards. Before we get to the problem of publishing data to the outside world, most data held by central and local government is not even shared properly within teams, within departments or within government as a whole. Once government is in a position to share data within itself, it is then only a relatively trivial step to sharing it with the outside world. The challenges around properly classifying data, 'marking it up', and annotating issues around source and quality, are not problems of open data *per se*. They are already addressed in many ways by the procedures of UK National Statistics. There are inherent problems around the exchangeability of data that need to be addressed before 'openness'.

The specific challenges around making data publicly available are around the resilience and scale of databases and webservers (which is a mundane though not cheap technical issue) and around whether any contractual interests attach to data that arises from government's interaction with the private sector (including pseudo-private bodies such as in the rail and utility industries).

Meaningful Open Data: Q1

Where public services document their processes, these processes should directly reference the data inventories. As a matter of sound data management, any data entering or leaving a department should appear in the inventory.

Meaningful Open Data: Q2

All data sets entering and leaving departments, and data directly appearing in reports (eg UK Statistics) should automatically be included.

Meaningful Open Data: Q4

The data inventories will be a first step in identifying data that is collected unnecessarily in the sense that it is duplicated. By mapping the data inventories to key high-level processes, it may also be possible to identify data that is collected but which appears not to be used.

Meaningful Open Data: Q5

Government should release the data that it holds and relies on, 'warts and all'. That seems almost axiomatic for an 'open government' approach. If there is reason to believe that the data is wrong, or if it is 'polished' before being acted upon, notes should be provided to this effect, along with an explanation of what steps, if any, are being taken to correct the issue. There may be circumstances where known quality issues are tolerable given the alternative cost of improvement/correction, but that should be an 'open' decision.

Government sets the Example: Q1

If the data is correctly stored, then the data can be stored once, and published by either means. If one has to choose one approach, then there should be a central portal, because this will encourage cross-departmental data analysis, and discourage the development of data siloes.

Government sets the Example: Q3

This should be the subject of consultation with identified interest groups, because the answer is likely to vary depending on the subject area.

Innovation with Open Data: Q1

There are already a large number of individuals and groups with the interest and know-how to do this analysis. Initially, government's role should probably be to publicise the existence of the various data sets to likely interested parties.

Robert Daniels-Dwyer MA (Cantab), MSc, PhD, MBCS

Enterprise Information Architect