



26 October 2011

Our ref: ICAEW Rep 102/11

Your ref: Open Data Consultation

Chartered Accountants' Hall
Moorgate Place
London
EC2R 6EA

Dear Sir / Madam

Open Data Consultation

ICAEW is pleased to respond to your request for comments on *Making Open Data Real: A Public Consultation*.

Please contact me should you wish to discuss any of the points raised in the attached response.

Yours sincerely

Kirstin Gillon

Technical Manager, IT Faculty



ICAEW REPRESENTATION

RESPONSE TO CABINET OFFICE CONSULTATION ON OPEN DATA

Memorandum of comment submitted in October 2011 by ICAEW, in response to the Cabinet Office consultation paper Making Open Data Real: A Public Consultation published in August 2011

Contents	Paragraph
Introduction	1
Who we are	2 - 3
Major points	4 - 8
Responses to specific questions	9 -50

INTRODUCTION

1. ICAEW welcomes the opportunity to comment on the consultation paper *Making Open Data Real: A Public Consultation* published by the Cabinet Office on 4 August 2011, a copy of which is available from this [link](#).

WHO WE ARE

2. ICAEW is a world-leading professional accountancy body. We operate under a Royal Charter which obliges us to work in the public interest. ICAEW's regulation of its members, in particular its responsibilities in respect of auditors, is overseen by the UK Financial Reporting Council. We provide leadership and practical support to over 136,000 member chartered accountants in more than 160 countries, working with governments, regulators and industry in order to ensure that the highest standards are maintained.
3. ICAEW members operate across a wide range of areas in business, practice and the public sector. They provide financial expertise and guidance based on the highest professional, technical and ethical standards. They are trained to provide clarity and apply rigour, and so help create long-term sustainable economic value.

MAJOR POINTS

4. IT has transformed the way that many businesses and industries operate, and has the potential to support a transformation in public services too. We welcome the government's efforts to promote open data and develop an effective environment for the proactive release of a wide range of public service data.
5. It is important to recognise, though, that open data is not an end in itself. It is a means to a range of social and economic benefits. While we understand that it is not possible to dictate how open data will be used, and it is important to let others find opportunities for using the data and developing new information services, it is also important not to lose sight of the ultimate objectives. These can help to frame decisions on priority, quality and funding of data releases.
6. We also believe that the government needs to put much greater emphasis on data quality, alongside availability. The benefits outlined by the consultation paper are dependent not only on data being available, but also on its quality. If data is inaccurate, incomparable or out of date, it will not enhance decision making but may in fact worsen it. It is also unlikely that businesses will be able to gain finance and build successful business models around open data if they do not have sufficient confidence in its robustness.
7. We recognise the desire to release data quickly and acknowledge that improving the quality of much public sector data is likely to be a long-term goal. However, we believe that there needs to be a far greater emphasis on improving data quality if open data is to achieve many of the benefits claimed for it. Therefore, alongside a strategy for opening up data, we recommend that the government develop a clearer vision for improving data quality and changing the culture in the public sector in this regard.
8. There are many aspects to improving data, from the processes around data collection and update to the governance and accountability around data quality. There are also many different types of public service data and the degree of quality demanded may vary according to the type of data and the context in which it is being used. Therefore, any approach needs to recognise the multi-faceted nature of the issues and encourage a range of actions. Crowdsourcing techniques may well have a role to play with regard to certain types of data. However, even crowdsourcing techniques require some infrastructure to succeed. All of these activities have significant resource implications at the very least, and we urge the government to consider in more detail how the accuracy, completeness and comparability of data can be improved to support the goals of open data.

RESPONSES TO SPECIFIC QUESTIONS/POINTS

Q1: Do the definitions of the key terms go far enough or too far?

9. We suggest that the government would benefit from greater clarity on the meanings of 'data', 'dataset' and 'information'. Although definitions are given in section 1, they are confusing and the terms seem to be used interchangeably in the consultation paper.
10. For example, unstructured factual data are included in the definition of 'dataset'. Unstructured factual data, however, are just that – unstructured. They are therefore not datasets, which we understand to be another term for structured files.
11. As a result, attention needs to be given to clarifying the circumstances when the government is really talking about datasets, in the sense of structured computer files/databases, and when it is really talking about any old data that might be held somewhere in the government's records and could in principle be extracted but only if they were searched sufficiently systematically and comprehensively.
12. The reason that this distinction is important is that the costs associated with making structured datasets open are likely to be much less than the costs of making relatively random data open. This makes a big difference to what can reasonably be expected of providers in different circumstances.
13. The consultation paper also highlights the distinction between 'data' and 'information', but then proceeds largely to use the two terms more or less inter-changeably. The distinction is important because there may be benefits *both* from:
 - 13.1. greater data transparency, as suggested in paragraph 7.8 of the consultation paper, if data are made available in standard formats (e.g. iXBRL) so that they can potentially be processed both within and outside of government in innovative and commercially productive ways;
 - 13.2. greater transparency of information, so that, for example as discussed in the consultation paper, government can be better held to account.
14. By confusing the distinction between data and information, the government risks confusing those implementing a policy of greater openness, with the potential result that the benefits of the policy would be sub-optimal.

Q2: Where a decision is being taken about whether to make a dataset open, what tests should be applied?

15. Paragraph 8.8 says: 'There is a strong need to bring about a behavioural and cultural change within public service providers leading to "open by default" being embraced. In the short-term there is a debate, for those datasets where quality may not be high, over how we should balance the immediate publication of data against seeking to improve quality. Given the costs of improving quality, our judgement is that we should publish data of lower quality in preference to holding it back, whilst seeking over time to drive up the quality of that data.'
16. There is a danger, in formulating governmental objectives in this way, though, that 'openness by default' will come to be seen just as an end in itself, rather than a means to a better quality of governmental information. This danger would be avoided if the objectives were formulated the other way round, as:
17. 'In the short-term there is a debate, for those datasets where quality may not be high, over how we should balance the immediate publication of data against seeking to improve quality.'
18. 'Given the costs of improving quality, such improvement will inevitably take a significant period of time. There is, however, a strong immediate need to bring about a behavioural and cultural change within public service providers, leading to "open by default" being fully embraced as the norm, while the overall improvement is taking place.'

19. 'Our judgement is therefore that we should immediately publish data, even if they are of lower quality, in preference to holding the data back. We should then seek over time to drive up the quality of the data'.

Q3: 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

20. Given that the processes and criteria for judgement are similar to those set out in paragraphs 6.6 – 6.12 of the consultation paper, there seems no reason why those requesting data solely for their own benefit or purposes should not be asked to pay the full economic costs of providing the data.

Q4: How do we get the right balance in relation to the range of organisations (providers of public services) our policy proposals apply to? What threshold would be appropriate to determine the range of public services in scope and what key criteria should inform this?

21. No comments.

Q5: What would be appropriate mechanisms to encourage or ensure publication of data by public service providers?

22. No comments.

Q6: How would we establish a stronger presumption in favour of publication than that which currently exists?

23. No comments.

Q7: 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?

24. No comments.

Q8: Are existing safeguards to protect personal data and privacy measures adequate to regulate the Open Data agenda?

25. While much public sector data contains no personal information and therefore can be released without concerns around privacy, this remains an important consideration. We recommend that the government adopts an approach that identifies privacy issues early in the process of releasing data to ensure that appropriate safeguards are put in place. While anonymisation techniques are likely to be important in this regard, they are difficult to implement in practice, with on-going risks of re-identification. Therefore, undue reliance should not be placed on such techniques to resolve privacy concerns.

Q9: 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?

26. It would be helpful in this context to make clearer distinctions between different types of data as they may have different resource implications. In particular, obligations around structured data are likely to have less onerous resource implications than obligations around unstructured data, where an organisation may need to collect bits of information from many different sources.
27. It is also important to recognise the on-going burden of increasing access to data. In many cases, data will need to be updated on a regular basis and access should not be seen simply

as a one-off activity. Therefore, the costs of maintaining and updating information on a long-term basis need to be fully considered in this context.

Q10: How will we ensure that Open Data standards are embedded in new ICT contracts?

28. No comments.

Q11: What is the best way to achieve compliance on high and common standards to allow usability and interoperability?

29. No comments.

Q12: Is there a role for government to establish consistent standards for collecting user experience across public services?

30. No comments

Q13: Should we consider a scheme for accreditation of information intermediaries, and if so how might that best work?

31. No comments.

Q14: 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.

32. No comments.

Q15: 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?

33. No comments.

Q16: Would we need to have a sanctions framework to enforce a right to data?

34. No comments.

Q17: What other sectors would benefit from having a dedicated Sector Transparency Board?

35. No comments.

Q18: How should public services make use of data inventories? What is the optimal way to develop and operate this?

36. No comments.

Q19: How should data be prioritised for inclusion in an inventory? How is value to be established?

37. No comments.

Q20: In what areas would you expect government to collect and publish data routinely?

38. No comments.

Q21: What data is collected ‘unnecessarily’? How should these datasets be identified? Should collection be stopped?

39. Data used infrequently can often be identified from the logged use of such data, in other words the recorded number of times that access is made to the relevant dataset. Often, this process is likely to be quite easy and a by-product of data management and normal operational

procedures. There may, however, be cases where such data, though used infrequent, would, when used, be crucial. The identification of unnecessary data is therefore partly a value judgement and should not be made solely by reference to the programmed analysis of operational logs.

40. Likewise, there may be cases where the data is clearly necessary to the performance of a governmental process or function, but that process or function is itself of no great value. The real question may ultimately be not about the data, but about the process. Open access to data will only improve government, or achieve the other objectives outlined in the consultation paper, if the accessibility of data continues to be considered in the wider context of improving governmental processes as a whole.
41. Where data can be identified clearly and unequivocally as unnecessary, it seems obvious that collection of that data should be stopped.

Q22: 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?

42. High quality is essential to the government achieving its objectives with open data. For anyone making decisions based on open data, they need to have confidence in the accuracy, integrity and timeliness of the data. If businesses are to build business models on open data, they also need to have confidence in the quality of the data.
43. Indeed, many of the concerns highlighted in the consultation document refer to quality of data, rather availability of data. For example, section 4.7 states 'The quality of data that is currently published is often poor, and publication may be intermittent, which is unhelpful for business in particular. Standards do not exist across departments or wider public bodies, so it is difficult to make comparisons. Data may be published without clear explanations of context, meaning that in reality it is difficult to use.'
44. The importance of quality is also recognised throughout the document, and we strongly agree with the comment that 'better data actually means less but higher quality data' (section 3.2). However, the approach outlined in the consultation appears to emphasise the release of data rather than the release of high quality data.
45. Putting the emphasis on the publication of data that are inherently flawed in quality, lacking in contextual richness and difficult to use is simply a recipe for flawed decision-making by those who use those data. If this includes businesses, then the outcome is likely to be to impede their progress and development rather than to assist them. In the worst case, publication of such data could be disastrous for small businesses, particularly start-ups, if it has the effect of sending them off in the wrong direction or exaggerating particular opportunities for development. Putting priority on the publication of such data for the sake of demonstrating the principle of openness could therefore sometimes lead, not just to unintended consequences, but to consequences that are the very opposite of those intended
46. We fully understand the desire to release data quickly, rather than waiting until the quality of the data is improved. Improving the quality of data, in many cases, is likely to be a time consuming and lengthy process. However, we recommend that, alongside a strategy for opening up data and changing the public sector culture into one of great transparency, the government gives further consideration to the issue of quality. This includes a variety of approaches for improving the quality of data, and changing culture to support this. It also includes the level of quality of data below which the publication of those data might lead to more harm than good.

Q23: 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?

47. No comments.

Q24: What factors should inform prioritisation of datasets for publication, at national, local or sector level?

48. No comments.

Q25: Which is more important: for government to prioritise publishing a broader set of data, or existing data at a more detailed level?

49. No comments.

Q26: 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?

50. No comments.

Copyright © ICAEW 2011
All rights reserved.

This document may be reproduced without specific permission, in whole or part, free of charge and in any format or medium, subject to the conditions that:

- it is appropriately attributed, replicated accurately and is not used in a misleading context;
- the source of the extract or document is acknowledged and the title and ICAEW reference number are quoted.

Where third-party copyright material has been identified application for permission must be made to the copyright holder.

icaew.com