

Rt Hon Francis Maude MP
Minister for the Cabinet Office

14th June, 2010

Dear Francis

TRANSPARENCY AND OPEN DATA: NEXT STEPS

1. Tim Berners-Lee and I are excited by the prospect of working with you and the Public Sector Transparency Board. We believe greater transparency and the opening up of public sector data are fundamental, and we are delighted that this is a key priority for the Government. The Government's achievements in the last month – the clear and unequivocal commitments in the Coalition Agreement, the Prime Minister's letter and your own statements and interviews, and the quick release of key datasets such as COINS – have been seen in the open data community here and round the world as major developments which has put the UK in the forefront of this field.

2. In preparation for the first meeting of the Transparency Board, and to help you consider your priorities, I thought that it would be helpful to expand on our earlier conversation and indicate some of the areas which Tim and I would see as the key agenda for the Transparency Board in taking forward the Transparency initiative and achieving continuing success over the next year:

Policy and Culture

- (1) The public "Right to Data" is a vital part of the approach. It is important that the release of data is driven by what people want, by the formats and frequency in which they want it and by how they want to use it. We need to make that effective in practice as well as in theory, and to understand why the Freedom of Information Act has been relatively unsuccessful in opening up data for free re-use.

- (2) However even with law on their side an individual can struggle against the bureaucracy – as has certainly been seen with the Freedom of Information Act. So we need to help to drive culture change in Whitehall and the wider public services toward transparency more generally. Making data available without being asked is intrinsic to truly transparent government. The boldness of the Government's early releases – on the pay of the highest earners and on COINS – have certainly sent a wake-up signal; but our experience so far has been that it is impossible to underestimate the extent to which every data holder thinks

Intelligence, Agents, Multimedia Group

School of Electronics and Computer Science, University of Southampton, Highfield Campus, Southampton SO17 1BJ United Kingdom
Tel: +44 (0)23 8059 7682 Fax: +44 (0)23 8059 2783 www.ecs.soton.ac.uk/~nrs

themselves an exception or the number of reasons data holders can devise to block release. The Transparency Board must educate and challenge.

- (3) We must establish clear principles for public sector transparency and tests against which individual public bodies can be judged. We must ensure that individual data owners cannot act as judge and jury in their own cause on withholding data. I attach a note of the 'public data principles' which I suggested to you earlier; I think that an early decision of the Transparency Board should be to adopt something along these lines formally for application across government.
- (4) We need a settled model for "public task" in the provision of information. This is most apparent in the case of the 'information trading funds', but it is also to be seen in other public bodies' instincts to hug data or to charge again for information for which the taxpayer has already paid. The economic analysis, and the views we regularly hear from the business community themselves, are unequivocal: data must be released for free re-use so that the private sector can add new value and develop innovative new business services from government information. It will of course be important also to look at whether value-added charged information services from government trading funds and other public bodies are necessary and compete fairly with private sector value-added services (for instance do competitors have equal access to the raw data?)

Getting the data out

- (5) For the datasets already committed in the Prime Minister's letter and in the Coalition Agreement we need to agree the standards and definitions, police exceptions and redaction policies, and ensure effective implementation so that the result meets the spirit as well as the letter of the commitments.
- (6) In addition to these key "high interest" datasets we need to establish a constant flow from the greater stock of more run of the mill information, including management, performance, regulatory, service, historical and comparative data. There's a "long tail" of data, and quite obscure data can prove interesting and useful. Data.gov.uk currently has a lot of statistics and aggregate data but relatively little transaction-level data (data relating to cases, events, schedules, meetings, testimony, reports, grants and contracts). Nor do we have real-time information, for instance about transport, weather, environment etc. all of which allows the building of some really useful end-user applications.
- (7) Transparency of data is not just an issue about central government. Local Authorities, the National Health Service, schools and police all have useful data which should be public. There is important work to do here to ensure that the same standards and energy we are seeing in central government are embraced in local government and elsewhere. Indeed many of the uses of public data that we have seen proposed link data from different public services into a joined-up presentation to a local community. Currently the Department of Communities and Local Government have a separate initiative on local data. I am involved with that, but I have found that it is neither efficient nor effective to have a completely separate team on this. I strongly urge that all this work should be brought together under

your leadership and that of the Transparency Board and supported by an integrated team in the Cabinet Office.

- (8) Transparency principles need to be extended to those who operate public services on a franchised, regulated or subsidised basis. If the state is controlling a service to the public or is franchising or regulating its delivery the data about that activity should be treated as public data and made available. For example timetables and the real time running information for trains and buses are currently not available for free re-use – even though it would seem to be in the interests of the companies, of the government (which subsidises public transport) and of the environment to make it easier for people to find out when and where public transport is available and so become paying passengers.

Building a sustainable transparency "ecosystem"

- (9) We need further work to develop and consolidate data.gov.uk as the single online point of access for all UK public service datasets. We need to work with departments to make this part of their routine operations. It is essential that this work keeps its 'skunkworks' flavour and does not turn into a conventional big IT project. We also need to stamp out duplication of effort across the public sector. There are a number of similar initiatives planned in individual sectors or agencies, both within central government and in the wider public sector. Not only does this waste money and cause confusion, but it also makes it harder for the public to get at data without intimate knowledge of the structure of government. These – and their funding – need to be brought together under the Transparency Board and data.gov.uk.
- (10) When Tim and I started working with government we could not find a reliable inventory of what data government actually holds. That still does not exist. So we cannot measure the extent to which Government as a whole, or individual departments, are releasing their data. Even more importantly the public do not know what data they could request through the "Right to Data". We need to ensure that departments have a systematic public record of what data they hold, how it is produced and formatted, and how accurate it is. In the USA this is a legal requirement on federal agencies; in the UK it is not.
- (11) We need to support the development of licences and supporting policies to ensure that data released by all public bodies can be freely re-used and is interoperable with the internationally recognised Creative Commons model. There are a myriad of different terms, conditions and licences for information - chaotic in the public sector where each authority makes up its own rules. This is particularly evident when it comes to re-use of information.
- (12) A key Government objective is to realise significant economic benefits by enabling businesses and non-profit organisations to build innovative applications and websites using public data. So we need to engage with the communities of developers and of 'information entrepreneurs' to ensure that we are not only providing the data they want in the form they want it but also that the right supporting information and services are available to enable them to exploit the data quickly and efficiently.

- (13) We must promote and support the development and application of open, linked data standards for public data, including the development of appropriate skills in the public services. Linked Data refers to a set of best practices for publishing and connecting structured data on the Web. This is exactly what is needed for the data that Government is publishing and this is exactly why Tim and I have been promoting this approach and others (in particular the US) are now following. Once public data is published in this way it is much cheaper, efficient and simpler for people and organisations to use and exploit it. (We believe that there would also be internal gains in efficiency and effectiveness within the public sector itself through the use of linked data standards. It is these benefits which are driving the rapid adoption of the standards in, for instance, the pharmaceutical sector.)
- (14) Although the absence of standards at the moment should not delay quick action to release initial data, it will be important to move quickly to ensure that departments, agencies or council with data about the same subject release it in a standard format and using consistent definitions. This helps the development of generic applications that can take consistent data from different public services and present it either as a 'local' view or as a comparative view of different services. This should also apply to standardisable reporting or other data by government departments, including Annual Reports, performance reports, and statutory and Parliamentary reports. We are already getting requests for recommendations on some of these standards, illustrating that if we can establish standards quickly in the current vacuum then they may be readily adopted without too much mandation.

International

There is a wider context of course. We should work with the Government to engage with the leading experts internationally working on transparency, public data and standards, and to promote international liaison and global standards setting. Indeed we need to recognise that other countries have ambitions in this area and that the Government will be judged not only by how much it has done from where it started but also by how it compares to other countries' achievements. We believe that the British Government's statements and actions in its first month have raised the bar, and we look forward to working with you to help keep Britain leading the world in Transparency.

I understand that it might be a few weeks until the full Transparency Board is in place. Perhaps it would be helpful therefore for us to meet to discuss these points with you and your colleagues.

I am copying this letter to Tim Berners-Lee, Tom Steinberg, and to Andrew Stott (Cabinet Office).

Best regards



Professor Nigel Shadbolt FREng
Government Transparency and Open Data Advisor
nrs@ecs.soton.ac.uk

Intelligence, Agents, Multimedia Group

School of Electronics and Computer Science, University of Southampton, Highfield Campus, Southampton SO17 1BJ United Kingdom
Tel: +44 (0)23 8059 7682 Fax: +44 (0)23 8059 2783 www.ecs.soton.ac.uk/~nrs

DRAFT PUBLIC DATA PRINCIPLES FOR CONSIDERATION BY THE TRANSPARENCY BOARD

- **Public data will be published in reusable, machine-readable form** – publication alone is only part of transparency – the data needs to be reusable, and to make it reusable it needs to be machine-readable. At the moment a lot of Government information is locked into PDFs or other unprocessable formats.
- **Public data will be available and easy to find through a single easy to use online access point (<http://www.data.gov.uk/>)** – the public sector has a myriad of different websites, and search does not work well across them. It's important to have a well-known single point where people can find the data.
- **Public data will be published using open standards and following the recommendations of the World Wide Web Consortium** – your manifesto already recognises the importance of open, standardised formats. However to increase reusability and the ability to compare data it also means openness and standardisation of the content as well as the format.
- **Public data underlying the Government's own websites will be published in reusable form for others to use** – anything published on Government websites should be available as data for others to reuse. This is a real issue since some departments and agencies try to restrict this so that people have to come to their websites!
- **Public data will be released under an open licence which enables free reuse, including commercial reuse** - all data should be under the same easy to understand licence. The Creative Commons interoperable licence which we have had developed is available right now. But this principle needs to be extended across the public sector. And we have been surprised to find that information obtained under the Freedom of Information Act is not always released under a licence that allows free use and reuse – it should be.
- **Release data quickly, and then re-publish it in linked data form** – Linked data standards allow the most powerful and easiest re-use of data. However most existing internal public sector data is not in linked data form. Rather than delay any release of the data, our recommendation is to release it 'as is' as soon as possible, and then work to convert it to the better format.