



HOUSE OF COMMONS

SELECT COMMITTEES

Public Administration Select Committee

Oral evidence: Statistics and Open Data, HC 564

Tuesday 8 October 2013

Order by the House Of Commons to be published on 8 October 2013

Members present: Mr Bernard Jenkin (Chair), Kelvin Hopkins, Mr Steve Reed

Questions 1 - 79

Witnesses: **Professor Helen Margetts**, Director of Oxford Internet Institute, and **Dr Ben Worthy**, Lecturer in Politics at Birkbeck College, University of London, gave evidence.

Q1 Chair: First, may I apologise to you for the late-running of our Committee? We are very grateful to you for joining us. This is our first session on open data. Could you identify yourselves for the record, please?

Professor Margetts: I am Helen Margetts, Director of the Oxford Internet Institute.

Dr Worthy: Benjamin Worthy, a lecturer in politics from Birkbeck College.

Q2 Chair: Thank you very much for being with us. May I suggest that, if you feel we are not asking the right questions, you give us the information you think we need in your answers so that we are educated? Could each of you just describe to us what you think the term “open data” actually means?

Professor Margetts: “Open data” is an interesting term and it gets many terms bundled into it. In some ways, open data is a bit like the little white bunny of government. It means something nice, warm and good that will aid democracy. It is a confused term, though, because it implies that it is just opening data up for public use. In the loosest sense, that would be a term, but some people use it to mean open government; some people use it to mean big data—data that can be used for a particular kind of analysis.

The confusion comes from the fact that it basically has three aims, and you will see all these aims mentioned by different people—and even the same people. They are: greater transparency; improving government, as somehow open data will make government better;

and encouraging innovation and enterprise in the wider commercial world. That is where the confusion comes from.

Dr Worthy: You have Rufus Pollock from the Open Knowledge Foundation speaking later. It has a very good definition of open data: information that is reusable and accessible to all, and can be read by a machine. That is a very simple definition. Helen is exactly right in what she says. The problem is that it is conflated with lots of other ideas and there is an assumption that open data means open government, for example. There are quite a lot of assumptions built into an idea like that—that open data then automatically leads to transparency and accountability of government.

Q3 Chair: What is the difference between open data and open government?

Dr Worthy: Open data is about technology and making something machine-readable. Open government brings in more political ideas about making government accountable and answerable. For example, freedom of information is about accessing information. Open data is about re-using information and doing interesting and new things with it to gain new understandings. There is considerable overlap there, but they are not necessarily the same thing, not least because the Government can control what open data is given out.

Q4 Chair: How well does each of you think the Government understands this?

Professor Margetts: Not very well.

Q5 Chair: Thank you. Why was that so hard to say?

Professor Margetts: For the reason I gave. To criticise open data is a bit like saying you do not like babies: you are not supposed to say it. It is true that it has very noble aims, and most of the people who are great advocates of open data have the three noble aims that I mentioned. The point is that there is an idea that open data will somehow automatically lead to open government, and that is by no means the case.

Q6 Chair: Dr Worthy, do you think in government there is a sense that they just need to be in favour of it and it all will happen, whatever “it” is?

Dr Worthy: It is a bit like the idea of transparency. Transparency is just a good thing in all situations, and it is so deeply embedded as a good thing that it is difficult to argue about the nuances. What Helen says is exactly right. There are so many aims attached to open data. There are economic aims about hunting down waste and encouraging use by business. There are political aims about participation and transparency. There are social aims about empowering communities. There are so many, and one of the things I have been finding in my interviews is that local government, from its perspective, finds that central Government are switching between different aims and shifting the emphasis. They are not exactly sure what it is they are supposed to see, whom they are supposed to be targeting with this data and what effects should be coming out of it.

Q7 Kelvin Hopkins: You implied that open data means open government, but open data has to have a licence, as I understand it, and there is also closed data. The Government can choose if things are open or closed. Is that right?

Dr Worthy: Yes.

Professor Margetts: Yes, Government Departments have to release data, but they can to a large extent choose which data to release and which data not to release. They cannot release some data because it is personal data or if there would be strong ethical or legal barriers to them doing so.

Dr Worthy: That is one of the fundamental differences to bear in mind between open data and open government. If you compare open data with freedom of information, open data is still what the Government choose to release or emphasise. With freedom of information, you can ask for many different things, some of which the Government probably do not necessarily want you to ask for. There is more of a sense that the Government cannot control exactly what has been asked for with FOI. There is a distinction.

Q8 Chair: If a graduate brought to you the open data White Paper as a piece of work, what sort of marks would you give it?

Professor Margetts: I would give it quite good marks as a plan for what you are going to do but, in a way, I would find it very difficult to mark, because we do not have enough data about the data. I think I would be asking for a viva.

Dr Worthy: My first comment would be a comment I often put in essays: needs more focus. It needs a stronger line of argument to tell me what exactly the aims are, and more evidence to tell me what is going to happen and why.

Q9 Chair: What marks would you give the Shakespeare review?

Professor Margetts: I think the same. It was good, but both these things are lacking feedback loops that would somehow make these things happen. It is no good just putting data out there and somehow thinking it will float off into the stratosphere, make things better and achieve these three aims. You need some mechanism for evaluating whether that is happening. That is what is so difficult. We are trying to do some of that at the Oxford Internet Institute. We are doing a systematic analysis of the data.gov platform, but some of the information you need just is not there.

Dr Worthy: There are some very interesting ideas in the Shakespeare review. There is the need for a more coherent strategy, with its emphasis on citizens as users, and the need for a feedback loop is very important. I have mainly been working with local government, and one of their points is they want to see what is in it for them in terms of public authorities. A lot of the talk is about what will happen for citizens, businesses and others.

A selling point being missed is how this sort of data could really transform how policies are made and resources are allocated. If you could develop applications that traced down to the street level how money is spent, that could make a huge difference to how a local authority makes policy and decisions, and how it allocates resources. Homelessness is one area that springs to mind. The Government probably need to sell more strongly what is in it for a public authority, as well as helping the public and other groups.

Q10 Chair: The Government need to have a clearer idea of what they are attempting to achieve by open data. Is that it?

Professor Margetts: Yes. I do not know if you have looked at the data, but it is not immediately citizen-friendly.

Q11 Chair: I was going to come on to the quality of the data. What do you feel about the quality of the data that is being released?

Professor Margetts: It would not be very easy for the average citizen to use the actual data sets.

Chair: It is very inaccessible.

Professor Margetts: That is going to depend on the development of applications that allow citizens to do something with it. Data is lacking, but of the 300 applications that are mentioned on the site, I think only 20 of them are registered with iTunes. That is all we can find.

Q12 Chair: How applicable do you think the term “data dumping” is? How comfortable do you feel with that epithet being applied to the Government?

Professor Margetts: I would say it is not quite fair.

Chair: Some people say that, don't they?

Professor Margetts: Yes, there is a danger, but it is difficult to generalise. There are 14,000 data sets on the platform, so that is a lot of data sets. They are tremendously variable in their usefulness and in the extent to which they are being used, but it is difficult to say because we do not have that level of data about the individual data sets.

Q13 Chair: Can you summarise what three things they need to do to improve the quality of the data?

Professor Margetts: To start with, if they want them to be used by citizens, they will have to be citizen-focused applications rather than developer-focused.

Q14 Chair: Who will develop these applications?

Professor Margetts: That is the thing. They need to be incentivised to do it if they are going to be citizen-focused as opposed to geared at making money or whatever.

Q15 Chair: Is the expectation that people out there are going to start developing applications a little naive? How confident are you that applications will be developed on the basis of the Government's present strategy?

Professor Margetts: Commercial use will be made of quite a lot of this data, but there is no particular incentive for the developers who do it to say that they are using the data—they can certificate it, etc., but why would they? They might as well just use it, because it is like a free good. Secondly, the incentive is going to be to make money out of it, rather than to do things for citizens.

Dr Worthy: To give you three: it needs to be contextualised; it needs to be comparable; and, to an extent, it needs to have a narrative, as was mentioned in a number of my

interviews. Even data needs to tell some sort of story. People will innovate with this sort of information anyway, because that is just their nature.

Q16 Chair: In the area of public policy, unless you have a Dr Foster or a mySociety, the citizen is not going to be able to make any sense of this. What does the Government need to do to generate those applications and those intermediaries?

Professor Margetts: It needs to encourage and incentivise intermediaries with a citizen focus to make that sort of aim explicit and to think of ways of incentivising NGOs to do it.

Dr Worthy: Perhaps it needs to meet them halfway by, to an extent, standardising information. Looking from local government level at the spending data, often it is just a list of columns: the name of a company, the amount of money and the directorate that has given that money to it. It needs more there. People at local government level are very interested in localised information and amenity. The classics are council tax, amenities and services. If they can do more in that direction—link it to other information, bring it closer and make it easier to innovate with—that might make a difference.

Professor Margetts: In part it depends on what your aim is. There would be different things to do for different aims. This would be great, for example, if your aim were to make Government more transparent. If you want to make government better and improve it, this may not be the right sort of data. Government Departments hold all sorts of data as a result of their administrative operations that could be put to greater effect in making government better. That would not necessarily involve making it open. It may not be possible to make it open. It may be too personal; it may be ethically impossible to make it open. It is data for different aims.

Q17 Chair: Can we have an example? For example, patient records and medical statistics are personal data, but in aggregate are huge sources of intelligence about how to improve public health, drugs and treatments. However, that information is not available to universities or the research institutes of major companies. Is that the sort of data you are talking about?

Professor Margetts: Yes, that is the sort of data I am talking about. It does not just have to be health; it could be tax or benefits administration data.

Q18 Chair: How do we deal with that?

Professor Margetts: It may not be possible to make some of that data open, but we cannot necessarily expect open data to do the job of that data.

Q19 Chair: So what we are in fact saying is that open data includes a collection of data that might not be open to the public but needs to be open to other users on a protected basis.

Dr Worthy: Yes, there is a halfway house for some data in that you can give it to people with certain limitations and commitments to use the data in a certain way.

Chair: That would be in the public interest.

Dr Worthy: It would depend on what the data was being used for and who was using it—a journalist or an innovator who is using this in some way. That can be controlled. That is the halfway house between closed data and completely open data.

Q20 Mr Reed: How have open data initiatives transformed or changed public services?

Professor Margetts: They have not, for some of the reasons we have been talking about, because of the lack of feedback loops. If one or more of these 14,000 data sets contained some really important information about the difference between contract providers, service providers or policing services, it is not very clear what the feedback loops are for that information to get back to the Department that has produced the data. It is very much what Beth Noveck called “throwing data over the transom”. You throw it over, and then it floats off and it is good. Without the feedback loops, it would be very difficult to point to that.

Q21 Mr Reed: There are no examples you can point to where this has made, or helped, a transformation happen.

Dr Worthy: There are examples locally where you could see it could have some potential. Mastodon C was one innovation that led to changes in medicines, for example.

Q22 Mr Reed: What is it?

Dr Worthy: How much it costs to get particular medicines from various parts of the NHS. You could quickly search it and find that some places were paying much more for similar medicine. You can see how that could have huge potential. There was one very high profile example of citizen auditors taking on Barnet council because it wanted to contract out a great deal of its services. There are some examples. Using one innovative website, Openly Local, which allows you to look at all sorts of different local council spending, one journalist did a story on Southern Cross and how much different councils paid. You can see these things, and Helen makes a really important point here: this information on its own will not make a difference; it needs to be attached to accountability mechanisms. From local government’s view, you can see how open data could match with participatory budgeting, local referendums or the increased use of online consultation. On its own, though, a bit like freedom of information, it does not just unleash a wave of accountability and feedback. It needs political mechanisms into which it can feed and then be used.

Q23 Mr Reed: Comparative performance data would work as well. Are there any examples of it increasing accountability in that way? What examples are there of local public services or national public services moving in that direction? There are a lot of small-scale initiatives that may not be being tied up yet. It would be useful to know some of those.

Professor Margetts: We were talking just before we came in, and some of the open data initiatives there have been, like crime data, have definitely got a lot of interest and a lot of people looking at them. Crime data is hugely used, but the extent to which it is being used to feed back into policing services is much less clear, because people are maybe looking at crime data because they want to compare house prices with where they used to live, or where they might go, rather than it being fed back into public services.

Dr Worthy: The accountability that I found at local level seems to be quite low level or localised. It is on a particular controversy about a chief executive’s pay or use of money in

a certain policy area. It is often emphasised either by a hyper-local site or by local journalists, and there may be some use of this by local activists as well, but sometimes it is under the radar, as are a lot of things at local level.

Q24 Mr Reed: If there is not much of it happening, what are the main barriers to publishing more data in ways that are useable by citizens or others who have an interest?

Professor Margetts: One of the ways is the one I mentioned before—some of the most useful data is actually personal data, or data that it would be very difficult to open up. Obviously, the incentive for Departments and agencies is to open up the data that is easy to open up and that they do not mind opening up, rather than the data that is going to have a lot of barriers to them so doing. That needs tackling head on.

Dr Worthy: I am in the middle of conducting a survey of local authorities about open data, and I asked them about some of the barriers. One barrier is cultural. It depends how the authority works and where open data could fit in in its mode of decision making. I think it is about leadership. We talked at the beginning about the strategies that were not clear on exactly what was wanted. Some leaders may see this as something that they should just minimally comply with. There is also an issue of resources. Some local authorities are locked into particular contracts with technology and software, so they can only do certain things. It may also be an issue about skills: to really take advantage of this data, link it to other pieces of information; and to help with innovations, you need particular skills that not all authorities or public bodies may have ready access to. One solution is to make sure that public bodies do work closely with innovators who do have this level of skills.

Q25 Mr Reed: Do you recognise there would perhaps be a risk that some politicians would feel that, if you moved towards publishing data around street-level spend, as you mentioned earlier, that could create an impetus towards equalising spend per capita, rather than focusing spend on areas or households with greater levels of deprivation or social exclusion, who would be generally less articulate?

Dr Worthy: Looking at local government especially, I have found that although technology is often presented as a neutral good, it is extremely political. The local government spending data is very politicised. It is about local versus central Government. Some interviewees have spoken of how they feel that the spending data is presented in a particular way to make it look like local authorities are profligate wasters of money. Regarding the way in which you see it now, with the £500 spends, there are automatically words like “bar bills” and you see how much money was spent. The way you present the data can frame what you think. There is much less information in their spend data about what impact the money has and much more about the fact that it has been spent, which biases you to understand spending in a particular way. Whatever way you push it, there are risks that it can have all sorts of effects, not least because how people behave when given information is very nuanced and subjective. It is not simply that they get a piece of information and then do X.

Q26 Mr Reed: Pushing the same point, how do you overcome the fact that, if you did have street-level spend data, most people would find that less than average was spent on their household, because 80% of public spending is focused on 20% of households? How do you get an explanation along with the data so that it does not get misconstrued?

Professor Margetts: I am just agreeing with you. Why would that particular data help to make public services better? It would not.

Dr Worthy: There is only a certain amount that you could do. I think you could push out as much information as possible and as much context. Like Helen said earlier, you could also have accountability and consultation mechanisms to try to equalise this. I found the same problem with freedom of information. There was a concern that you sent out an FOI answer and some Departments and bodies tried to put as much context as possible into the answer to explain exactly what it meant. Whether the person receiving the information read or believed that context is another issue; it is very difficult to know.

Q27 Mr Reed: Can I imply from your responses there that you believe there is a limit to how far you think open data should go? Is there some data that, if you publish it, will not help anyway and might cause a problem?

Professor Margetts: Yes, but with the proviso that it might satisfy one of these three aims or principles, but not necessarily the one that you were aiming at satisfying. You have this dislocation between producers and consumers—somebody is producing the data for somebody else rather than for their own self-improvement.

Q28 Kelvin Hopkins: I am very concerned about local authorities and how data can be used. How could it be used to increase the understanding and public accountability of public services?

Dr Worthy: The main thing that comes out very strongly from the research I have done up until now—it has been running for a year—is that it is all about the applications and innovations that people have made. Very few people are going to go to their local authority website and look at the spending data, but many people are very interested in local amenities and local services. If you look at applications like Openly Local, it allows you at the click of a button to look deeply into one authority's spending or compare authority spending. You could also look at other applications I have mentioned in my evidence, like AppGov, which is also in preparation. First, if you can make it easier to understand, contextualise it and link it to other information and, secondly, if you can make it easier to access and understand, that can make a real difference. America is a good example here. There are lots of very interesting and radical experiments at the city level in America, where they have joined up contact centres with the real-time online ability to report problems and things, which is linked in to service data.

Q29 Kelvin Hopkins: Ensuring data is consistent and harmonised across local authorities is crucial to making comparisons. We have seen problems in the last few days in Birmingham, with a desperate shortage of social workers—they have admitted they cannot cope—and there are other local authorities doing well. It should not get to that stage. If we had good data and good comparisons between local authorities, perhaps we could intervene at an earlier stage and make sure these things do not happen. Is that not what some of the data should be for?

Dr Worthy: Absolutely. You could imagine that you could have an application that could measure the number of hours social workers are spending on average with different people in different Departments. It could work as an early warning system for this sort of issue. The problem is that these are also very political issues about social care. You may have

seen last week that Eric Pickles talked about publishing comparative car parking data for local authorities to see how much money local authorities made from car parking. It seems quite innocent, but of course it is a very political issue. If you spend a lot of time around local government, car parking is huge, and you could see how this could go in particular ways. You are exactly right, though, that that is how it could work.

Q30 Kelvin Hopkins: It strikes me that the abolition of the Audit Commission was not good because the commission has done a lot of this kind of work. Leaving local authorities to their own devices more, and perhaps even encouraging them to contract out, makes it perhaps more difficult to compare. Is it possible that Governments of either colour might not like some of this data? It could, for example, show that where you have direct services rather than contracted out services, the direct services are more efficient, produce a better service and are more accountable. The flavour of the last two Governments was that they wanted to get as much out into the private sector—into contracting—as possible.

Dr Worthy: That is exactly the point: even though this open data looks neutral, it has very political implications. One of the concerns I have been finding in my research is actually the opposite. The concern is that by publishing all this data about public service performance, you are handing potential ammunition to private providers who do not have to provide that sort of information. It enables them to point at particular public services and say, “Look how expensive X is. Look how wasteful Y is.” It is a very easy thing to do, so there is a concern. Helen spoke about a piece of research that described it as a potential Trojan horse for further privatisation.

Q31 Kelvin Hopkins: Would it not be sensible for the Government to require the private companies to provide the same information that local authorities are required to provide? We could have genuinely open data across the board.

Dr Worthy: Absolutely.

Q32 Chair: Excellent. That has been very informative. Is there anything you want to add?

Professor Margetts: You are going to be doing a lot more questioning and you are going to be talking to—

Q33 Chair: Tell us what we should be asking our other witnesses.

Professor Margetts: You should ask them about providing more data about what is happening to the data, and feedback loops between the producers of the data and the users of the data, so that the data gets better.

Q34 Chair: Do you think the punter or the ordinary member of the public—one has to be careful about how one says that because there is no such thing—wants more data that is specific to the question they are asking, or that gives a picture, story or narrative about the subject they are looking at? What do you think we really want? Is it just about the death rate in my hospital for patients concerning heart disease, or do I want to understand what that means in the context of all hospitals, all death rates and all heart disease throughout the

country? Is data meaningful on its own without context? Sorry, I have asked about 10 questions.

Professor Margetts: In an ideal world, the idea would be that you have some kind of intelligence centre that can put this data together and produce conclusions from the data as a whole, rather than individual data sets released by individual health trusts and individual local authorities. The ideal—the kind of thing that Tesco or Walmart has—is some kind of intelligence centre using data from all the devolved bits of the organisation. In an ideal world, government would be able to do that too.

Dr Worthy: I have just a few thoughts very quickly. Firstly, the politics of this thing is very important to bear in mind. A lot of the discourse about technology is very deterministic—“it is a good thing”—and perhaps you need to bear in mind the politics. There are a lot of assumptions about how people behave when given information or what people are interested in. The discussion about context and localising information is important.

I would just like to emphasise how much potential there is here, particularly in the innovations that can be developed. I teach a course about Parliament, and a website like TheyWorkForYou absolutely transforms how we can understand what Parliamentarians do. Other developments like the Open Knowledge Foundation’s “Where Does My Money Go” have huge potential to transform our understanding of how money is spent and what it is spent on. You need to bear in mind that there is huge potential to transform how services are delivered and how they can be understood—if only the data and the technology are used and understood in the correct way.

Q35 Kelvin Hopkins: Just to reinforce what the Chairman was saying about hospital death rates, there was a substantial debate on the radio last week, and many people challenged whether the particular statistic that is used to compare hospitals had any validity at all, because all sorts of factors are not taken into account. Is it important to put these qualifications in with the open data so that people do have a degree of caution when they are looking at them?

Dr Worthy: Absolutely.

Q36 Mr Reed: On that example you gave, I am a relatively new MP and I discovered that, on TheyWorkForYou, if you do two interventions in a debate, you get written up as having spoken in the debate, which is a tick, and it can take you half an hour. You could spend a whole day visiting five different community organisations and get nothing. Is that helping, or is it making MPs do things in order to get ticks on that website that are not necessarily helping them represent their electors?

Chair: Or, indeed, you can sit all day in the Chamber hoping to be called, having spent hours working on your speech, and then you never get called.

Dr Worthy: There is an interesting question about what sort of information is displayed, and I know Tom Steinberg is going to speak next. If you look at the research in 2011 on TheyWorkForYou, you do find, for example, that the site is attracting a huge number of people who previously said they were not engaged in politics. It does promote

understanding, although what should be contained in there is obviously tied into the idea of what an MP should do, which is a huge issue.

Chair: That is for another day. Thank you very much indeed. You have been really helpful. If you want to stay for the other evidence session we have now, that would be marvellous.

Witnesses: **Tom Steinberg**, Director, MySociety, **Heather Savory**, Chair, Open Data User Group, and **Dr Rufus Pollock**, CEO, Open Knowledge Foundation, gave evidence.

Q37 Chair: Please identify yourselves for the record.

Heather Savory: I am Heather Savory and I chair the Government's Open Data User Group.

Q38 Chair: Can you just remind us what the Open Data User Group is?

Heather Savory: We are an independent advisory body to Government. Members are appointed by me and the Minister for the Cabinet Office. The members are all volunteers and come from all areas of the data community.

Q39 Chair: You bring a particular perspective, being a formal career civil servant.

Heather Savory: My background is that I used to be in high tech, and then I did various other things. I am not a career civil servant, but I had two years in Treasury and three years in the Department for Business, Innovation and Skills.

Tom Steinberg: I am Tom Steinberg; I am the Director of mySociety.

Dr Pollock: I am Dr Rufus Pollock. I am CEO of the Open Knowledge Foundation, which is a non-profit promoting access to information and its use to improve the world.

Q40 Chair: Each of you is very interested in the open data agenda. For each of you and the people you represent, this could be a really big thing.

Dr Pollock: Yes.

Chair: How much do you feel the Government have really gripped this agenda and are taking it forward in the way that you wish?

Dr Pollock: Our first point is that the Government here in the UK have been a leader globally in this agenda. This Government and even the previous Government took serious steps to open up information and, to some extent, set policy. There are some obvious lacunae in that, but on the whole I think the Government have been leading. There is a question at the present time—we do not want the Government to rest on their laurels. While there have been some very good achievements, there is still much to be done. Most fundamentally, one of the big lacks—Tom was on, and I am still on, the Transparency Board—is the challenge you see in many countries with open data agendas: it still depends in some sense on Executive whim, to put it in its most blunt sense. If the Prime Minister or the relevant Minister were to decide they were no longer interested in this, there is no embedding of this agenda in legislation beyond what exists in FOI or things like that.

One of the greatest challenges is the certainty and the fact that, for example, you do not have this embedded in procurement rules. You do not have it whenever the Government buy a new IT system or every time they decide to outsource a service. It should still be the case that relevant data is still controlled by the Government. One of the biggest risks and dangers we have seen evidence of both in the US and here is that you outsource some service and, bam, all your information is gone. One example has been train timetables, where one of the challenges has been whether the Government actually control train timetables anymore, because obviously there are now private train-operating companies. You might think that train timetables are a fundamental piece of Government information or institutional information.

There has been some very significant progress, and the amount of data the UK Government have released has been quite impressive and it is probably up there with any other Government in the world. However, there are some things missing in the strategy and the policy. Most fundamentally, that is the embedding of this agenda in something that you can be guaranteed will live on. Obviously, you can never guarantee legislation will not be reversed, but there could be greater certainty that it does not depend on the current set of Ministers.

Heather Savory: I am the new girl on the block because I came to this agenda about one year ago. I have to say that I do not have all the history, but I think that the Government are doing quite well. The problem is that what we see, and what the Open Data User Group is here to do, is the mismatch between the data sets the Government want to put out and the ones that have the most economic benefit.

In the year that I have been on this agenda, we have seen some quite good improvements in the use of open data by the public sector. The area which I think we really need to focus on moving forward is to ease the commercial use of open data for the benefit of economic growth. By that, I do not just mean start-ups and apps; I mean the whole of the digital economy, which has considerable worth. A report has just been commissioned by Oxera that says this market is going to be worth \$270 billion this year and it grows at 30%. I have a passion for the UK growth in technology, and open data is a fundamental underpinning component of the UK being able to do something in this space for growth—to grow some new companies and to help existing companies.

My short answer is that the Government are doing okay, but they really need to focus on the data sets, and the fundamental core reference data that will underpin the value of those data sets, to make them usable by the business community.

Tom Steinberg: We have had an optimistic view and a neutral view, so I might as well be the pessimistic view. I think the Government have not shown willingness to be serious about the economic benefit, or the public services reform and improvement benefit. In fact, I feel that those of us—myself included—who represent this generation of people trying to reform Government are a pale shadow of my parents' generation in the UK and, on the other side of my family, my grandparents' generation. In both cases, there were some really substantial legislative changes out of the US and UK Governments in relation to the release of information. We are nowhere near there at this point in time. That is best shown by the fact that all the most economically significant data that Heather was alluding

to before is still only made available in ways that seem highly likely to me to be sub-optimal for the economy, simply because releasing it would be so politically painful and have some short-term costs to a Government who do not have any money.

On the political side, and the Government accountability, performance and service delivery side, which is where I have slightly more passion than the economic side, we have not seen any move forward regarding the information I can get to find out if my local council is conducting itself well, efficiently, or making terribly bad decisions that will affect me in my life. I feel that the enthusiasm that has been shown has come from a pretty small part of the Government—basically one or two Ministers—and the resistance from all the other Ministers has essentially meant that this Government cannot be the same as Governments before, in the '60s in America and in the '90s in the UK, in terms of making meaningful legislative change of the kind that means we will see more substantial improvements come to this country in economic and political terms because of this agenda.

Q41 Chair: How much do you feel the Government have not set the right priorities in this programme? What would be the right priorities?

Dr Pollock: Partially, to complement Tom—

Chair: By all means debate with each other.

Dr Pollock: Tom is right to be harsh. I think at the same time I possibly have a sense of the challenges we mentioned. Ultimately, one is that we have existed in a period of austerity. Frankly, one challenge has been the Treasury. Let's be concrete. The classic example is the postcode database, now interlocked with the privatisation of Royal Mail. There have been other key geographic data sets, and a lot of the barrier to some of this very important data, both for accountability purposes and commercially and economically, has been the fact that there is some “give up some jam today”—often a very small amount of jam—“for jam tomorrow”. Essentially, you might have to put up £20 million, £10 million or £40 million for what seem to be very significant anticipated benefits. The sense is it has been a challenge within Government to make that case to some Departments. Heather has done some sterling work but, for example, I do not think we have been totally successful on the postcode address file. In a sense it has been quite a group of people—all the Transparency Board members, for example, including Sir Tim Berners-Lee—saying that the Government should make that available. It would have very significant benefits, but it has been a real challenge. That is where it has been difficult, and Tom is right to say so.

At the same time, I would say I have seen some Damascene conversions among particular Ministers, in the sense of seeing the power of open data for them. I noticed one of the questions was also about the Civil Service or public sector seeing this. I do not think we could say that every civil servant today suddenly sees this, but I have seen certain Departments that have basically gone from hostility to this idea to significant endorsement and seeing a value to them in terms of quality, public support and trust—and other things. At the same time, it has been mixed. I do not know if Tom might be more critical on this view.

Heather Savory: I just wanted to come in and say I support what Rufus is saying. There is a particular issue with some of the trading funds. That is because they are last century's

model. They have essentially a funding model. They are set up as bodies that are required to self-fund through commercial activity. I was introduced to the four main trading funds here: Companies House, the Land Registry, Ordnance Survey and the Met Office. I refer to them as four non-identical twins. They are not all the same; they all do very different specialised work. When you get under the skin of this, what you see is that the Land Registry has a very good, very proactive open data strategy. We were able to persuade them to release the historic price-paid data for housing and we have recently been able to persuade them to release something called the INSPIRE database, which is boundaries for properties. The Met Office has a good open data programme and Ordnance Survey also has a very successful open data programme, but it does not go far enough if you think about it in the wider economic terms in which we should now be thinking about it.

They have a funding model; the Treasury gets a small amount of return. The Government have to pay for weather services and for geographic services. These are essential components of our society, and also there are security issues. There has to be a firm, underpinning set of data that we can operate the country on. The model there needs to be revisited in terms of what the wider economic benefits are. Rufus is right that there has been progress, but I think more needs to be done in some of those areas.

Q42 Chair: We are going to be writing a report on the back of this inquiry and we are able to make some recommendations. Can you briefly say what you think our recommendations should be?

Heather Savory: The absolute first priority from the data community, but also, since I have been looking at it, for the wider economy, is an open national address data set. There is a data set that is currently owned by GeoPlace, a joint venture between Ordnance Survey and local authorities, so that is in public ownership. Our fundamental push has been to get this released as open data. It is essential. In the same way that the internet is the infrastructure of infrastructures in the virtual and digital world, this geographic data is the infrastructure of infrastructures for everything else.

If you think of energy being a core infrastructure, and it is—energy is core to the UK—to supply energy you need to know where it has to go and how it is going to get there. If you do not know where you are individually, you do not know where you are going. That is the absolute, fundamental key to open that up. That will then underpin deriving the value from the very many other data sets that the Government holds in environment, transport, etc., because it is the connective tissue, as Stephan Shakespeare calls it in his report, that enables you to put data sets together because you get a common identifier. That is the absolute priority.

Chair: Right, that is one recommendation.

Tom Steinberg: There are two things from me that both relate to the fact that the discussion in the last few years has been much too much about whether we can get this or that data set out. The laws that this Parliament passes do not ever tend to be about really specific things; they are not about one person or one company. They tend to be about classes of companies and classes of people. We need two kinds of new law that do not exist at the moment. One is about government procurement of computer systems that simply says all parts of government, whether large or small, must not or cannot procure

anything digital or electronic of any kind that does not make the production of open data essentially available at zero marginal cost beyond the contract value. The reason that should be there is because it does not cost anything else. It is a bit like if you are building a brand new house and you want to make it easy for someone disabled to live in: it is incredibly cheap to build it in at first compared with taking a building like this and trying to retrofit it.

Q43 Chair: What is your other recommendation?

Tom Steinberg: The other one is about expanding the Freedom of Information Act so that people have similar powers to access data sets to those they have to access paper documents.

Q44 Chair: Presumably being able to access them electronically would make FOI much cheaper.

Tom Steinberg: It could do. The procurement piece—the first of the two pieces of legislation I mentioned—is what would really make it cheaper, because then if an officer or someone was asked for some data by a citizen or a member of the public, it would be incredibly cheap and trivially easy when they had to go and get that information.

Q45 Chair: Just finally on that, what is your recommendation or recommendations?

Dr Pollock: Those are good. There are a couple of things that I would add. One is that you would like to have the principles set out that all non-personal data should be open by default. Heather is doing a sterling job, but it was rather interesting that she had to go to persuade the Land Registry to do this or persuade X to do that. It should be the case that there are exceptions to the rule and those have to be strongly justified on some grounds. The point that Tom is getting at is that in theory things like that are in FOI, but they should be operationalised more effectively. Often it is so costly to do this that we cannot do it, which is the reason why we want the one-click button for open data in procurement rules. The open-by-default principle should be enunciated, and it should also be applied to the trading funds. That is one of the primary things I would request. The other is that something be put into legislation for these things. That is important.

I know this is subtle, but it is related to FOI, because a lot of it has gone into modifying FOI. The one thing Tom did not mention is that we also want to make sure that that data is made open. If data is provided under FOI or other stuff, it is automatically licensed under the open government licence. Otherwise, there is this silly thing where you get the data, but you are not allowed to give it to anyone else without permission. There is this weird thing where you get something under FOI and you are not allowed to give it to others, which is bizarre.

Q46 Mr Reed: Tom, just on the point you raised about open data, I used to be leader of a council before I was in here, and we implemented an open-by-default policy. When we started putting the data out, people started coming back saying, “Great, but your data is in the wrong format. We can’t use it.” It was a slow and expensive process to try to get that right. What is the process for making sure the format is going to be useable as universally as possible?

Tom Steinberg: First, I mentioned the need to put the requirement in procurement rules, because if it is not in procurement rules at all, it is very unlikely that the computer system that you then acquire will make it easy. After it is written down on the statute book, the structures needed to work out what standards are required are just general digital standards committees. They are not the most exciting sounding things in the world, but there are lots of them that basically make the internet work. Very few of them are specifically British, because the questions of how to send a picture, an audio file, an e-mail or whatever are completely global issues. In a very similar way, how I would release the single location of a bus stop is a completely global issue. The standard for that particular thing I just mentioned largely came out of America, but is an internationally negotiated standard.

The good news for the Government is that you do not really have to set up anything. You do not have to go and say, “Now we need our own standards committees.” Instead, all you do is find the people who are already working on the standards around parking places or whatever council information it is that people want, and you just participate in those. At the end of it they will say, “This is the standard. This is the one you should be using at the moment,” and then you make sure any new computer system you buy is compliant with that. You do not have to reinvent this wheel substantially yourself, because people all over the world are working on these common issues right now.

Dr Pollock: To add just one thing to that, the definition of open data at opendefinition.org does include machine-readability as a principle. You always have those stories in movies where the plucky defence are trying to sue this big company and they get given 500 box files of paper to read through. There is a way to prevent the use of open data: by giving people material in a way they cannot process very easily. Paper is not machine-readable. A computer cannot read your handwriting very well yet, and all these kinds of things. That is a requirement.

The key thing to say is that once you have given out something machine-readable—and by that I mean a spreadsheet versus a document with a table written on it—do not get too worried about the format. While geeks will complain and they will say, “You gave me Excel when you should have given me something else,” or whatever, in general once it is in that structure, people who are motivated enough will be able to use it or convert it. The one addendum to that is when you are missing some essential piece of information. A classic example, which is where it gets more subtle at the moment, is that the Government give out very detailed spending information now. They now give out contract information. You might think you would like to put those two bits of data together more effectively. Unfortunately, in the spending information, you do not tend to get the contract number that that spending relates to, if there is a contract.

That is where we get a little bit subtle, but often missing that key piece of information makes it very difficult to join up data sets that would give you insight through that joining up. Heather talks about connective tissue—addresses or postcodes: every citizen in the UK knows their postcode. However, if I said to you, “Do you know your longitude and latitude?” you will not. Those are often key moments where people, for example, know their geographic information. There are so many services you could use, such as when you go to Amazon or when you go to pay your gas bill. There are so many things that use your postcode; it is a connection between so many datasets, because it gives your geographic

information, which then allows you to connect into other data sets. That is a key point that is difficult in general to legislate. The key point is that we want to be able to connect data together. That is a key point: the power of open data is that it enables us to connect things together and therefore to gain better insight by joining up sets of information.

Q47 Mr Reed: Would there be conventions around that as well?

Dr Pollock: Yes, there would.

Q48 Mr Reed: Following on from what you said, it would be interesting to know the total public spend for a particular geographic location, but that is a near impossibility. Other problems are how granulated data is or how different data sets are aggregated together. Are there conventions of the kind Tom referred to we could tap into around that, or do the Government need to guide development of conventions of that sort?

Tom Steinberg: Many of these things already exist. You just mentioned a wonderful tool: you mentioned a thing that does not exist yet, which is how much the public sector spent in 100 square yards on different services. It is a lovely idea, and I am sure many of you can see how you might use it yourselves as politicians. It is a new idea, though, and because it is a new idea, you will probably find that the data standards to support it do not exist yet. I would say the Government absolutely have a role in participating in those decision-making exercises, but if I were going to ask how you put such a nebulous idea into law, I would say it is relatively clear. If a computer system is being provided in an area, that bit of government that is buying the system in that area must use the standard that is being internationally agreed right now. If there is no such standard, it should lobby for one and it should participate in the creation of one. That is relatively simple idea: if you are in a field, go and find a standard in that field, use it, and do not make your own up if there is already one that someone else has. That does not strike me as too difficult to describe.

Heather Savory: There are several questions here. As Tom rightly said, there are ubiquitous digital standards for data ranging from CSV to Excel data—as in normal Excel spreadsheets—and up to linked data. You may know that there is considerable work being done at the Web 3.0 Consortium about linked data. The premise is that the tech community can cope. You do not want to be re-inventing or prescribing specific things. There are preferred standards; the preference is to deliver things in a machine-readable format that technical people can cope with. I would also like to say that there are now lots of cheap or free platforms on which you can do data analysis using free software. It is not just the highly technical people who can deal with these data sets. One of the things we push for is more APIs. An API is something that allows you to pull out your own set of data from a big set of data. You can say, “I want this, this, this and this,” and you can extract it out. Whilst these standards all exist and there is a plethora of different things being delivered, the primary thing is that a document or a PDF is not a piece of open data. We have now got that far—that is not open data.

Across Government and all public sector bodies, we need some more help—not to say, “You must do X, Y and Z,” but giving us another way that you can get data out. One of the things that happens at the moment because of the silo nature of Departments and public bodies is that I expect there is some committee somewhere and then some poor chap sitting down in the basement—because IT people never have windows—and somebody will come

knocking on the door and say, “You’ve got to deliver that as open data.” This poor guy will be going, “I wonder what that means. What do I do?” Then whoever it is will go back upstairs and carry on with their policy job, because we have told him this organisation is going to be delivering open data.

Some technical guidance would be useful across Whitehall to give people more assistance about the preferred ways and challenges of delivering this open data. Some people will find it easier than others. I fully support the changes to the procurement rules because, as Tom has said, building in the method of delivering this data does put it at a real marginal cost. You might be astonished that, with current systems, our data is sometimes locked up and some of the incumbent IT suppliers to Government charge us for getting that data, because they have to extract it and nobody else can do that. That is fundamental to moving the agenda forward and making it work.

The other thing that I would just point the Committee to is that the Open Data Institute has just released a certificate—a method where you can go and self-rate your own open data. It is in beta; you can find it on its website. That is a really useful tool for the community because it has all the bits that you need to know. I would say it is a little bit fluffy in terms of supply and release dates, because if I am going to build a business based on this data, I want to know that it is going to be maintained. It is no good having a data set now, updating it in six months and then leaving it for three years, because my business will be dependent on it. This shows you where you can get it and it gives a good indicator of the quality of the data, so I think it would be a very positive move to have that and some sort of help or guidance note that went across the picture.

There is one technical point I wanted to bring up in legislation terms. I have to refer to my notes here; I have to get it right. There is the Freedom of Information Act and there was a Protection of Freedoms Act 2012. This is regulation that has just been brought into force, and this established an enhanced right to data because it introduces a statutory duty for public authorities to publish their data for re-use. I absolutely agree we do not have clean legislation. There is not an Open Data Act, but if you look at the complex network of legislation that we have, there is already a presumption to publish data and there are duties for public bodies to make their data available for re-use.

Q49 Chair: That is very useful. Very briefly, before we move on, how confident are you that Government have the right skills to do open data?

Dr Pollock: In some ways I do not think it has to be super-complicated. Some of it is reasonably straightforward in the sense of, “Publish it.” I would go back to this point that you can adopt best practice and so on. If the Government are doing machine-readable, people will complain and they will bug you, but if we are getting it out there, you are doing okay. Government obviously has a reasonable amount of sophisticated knowledge, certainly in IT teams and so on, so I do not think that is a barrier right now. There could of course be more knowledge about data, and it is also a generational aspect, but a lack of IT knowledge is not what is preventing the Government from passing legislation right now.

Q50 Chair: Does everybody agree with that?

Heather Savory: I agree with that. I think there are awareness and attitude problems.

Q51 Chair: I was going to ask: what is the attitude problem?

Heather Savory: There are perceived risks among civil servants, which is natural because they have been brought up in a world where they are protective—they are protective of the public and they also have concerns about their individual responsibilities. Many times I get re-quoted at me the poor chap in HMRC who left the CD on the train as the reason we should not have open data. They have genuine concerns about the privacy of the public, because privacy does need to be maintained. They are also concerned because their data is not perfect, but no data set is perfect. So there are those cultural issues that we need to find ways to give people confidence about. There is also a lack of belief that the technical community can deal with this stuff, whereas there are some workshops that we have held where is a guy who said, “Just give me big, dirty data. I’ll deal with it.” That is a place where those organisations can add value.

Q52 Chair: How do you think the problem of attitude and awareness should be addressed?

Heather Savory: I think it is difficult. You could promote the benefits to the economy, promote the growth agenda and show people case studies of how data can be better used. There are lots of examples now of where open data is used in the delivery of public services.

Q53 Chair: So they need to be made aware of those examples.

This is a question for all of you: how well do you think the different bits of government are working together on open data, such as the Cabinet Office, BIS and other public bodies?

Dr Pollock: You also asked a similar question. This is almost being a little bit facetious. There are Civil Service awards and so on, and one thing that was debated was that you could have a big open data prize and say, “This Department”—or this civil servant—“last year did the most prestigious thing,” or things like that.

Chair: We do not usually make recommendations to *Civil Service World*, but we could add it to the list.

Dr Pollock: Exactly. You can ask the base question of, “How did your attitude change?” In a system like this, it is partly a slow process, but it is also done by creating champions and by highlighting people who have been pioneering. The point Heather is getting at fundamentally is that the risk/reward setup of Government tends to be risk averse.

Q54 Chair: What about how Departments work together on open data?

Heather Savory: Do you mean in terms of driving the agenda? Obviously you have the Cabinet Office Public Sector Transparency Board and you have BIS, with information economy, growth and the Public Data Group, which is the four trading funds. Then you have the MOJ and National Archives, which are doing all the legislative side, and then you have the Information Commissioner’s Office. Do they work together? Yes. Do they work together well? Most of the time. Are you going to be able to change that at policy level? No, because you need all those elements of the policy.

Chair: This is becoming a pet of mine: it is the usual “headquarters of Government” problem. There is not one.

Heather Savory: There are always tensions there.

Q55 Chair: You are so diplomatic—much more than me. Do you have a prize for the best Department or the worst Department?

Heather Savory: The Office for National Statistics.

Chair: Is it the best?

Heather Savory: Yes.

Q56 Chair: That is nice for us to hear. Who is worst?

Heather Savory: It is difficult to say who is worst. The ones that I find the most irritating are those who release data sets and pretend that they are under the open government licence and then you discover that they are not under the open government licence. It is a bit of a pseudo-open delivery of data.

Q57 Chair: Can you just explain that? If it is not on the open government licence, why is it not open government?

Heather Savory: That would be my question to them. There are a lot of published data sets that, until you actually go and look at them, you think you are going to find under the open government licence, and then you find they are not or they are not quite.

Q58 Chair: What does that mean?

Dr Pollock: That they are not open.

Heather Savory: It means they are not fully open for use and re-use. It means they are not free. It means that there are licensing conditions attached to them.

Q59 Chair: And this affects the potential business opportunities.

Heather Savory: Absolutely, yes. You have to be able to let this data flow through the system.

Tom Steinberg: I have an answer, which is it is the Members of the House of Commons and Lords who have, as of yet, not updated our information legislation so that, as more and more bits of the public sector get outsourced, we will gradually lose sight of essentially how government works. Our children will know much less about how the Government work and whether or not they are doing a good job than we do. This is a relatively well-known issue, but I have not yet seen it become a really big enough issue that anyone has tabled a serious amendment. It is an absolutely terribly serious issue.

Q60 Chair: Is that in your written evidence?

Tom Steinberg: I am afraid I have not provided any written evidence.

Q61 Chair: Can you send us a note about what you really mean by that so that we have something concrete that we can refer to in our report? I would be very grateful for that. Blog it if you wish. Dr Pollock?

Dr Pollock: The Office for National Statistics has done well. Some of the Departments that are the parents of trading funds have done less well, but that has now been centralised to some extent into BIS. There has not always been great oversight of the trading funds by their parent Departments, perhaps naturally so. Now it is the Shareholder Executive plus a parent Department. Recently, there has been this inventory where the Government are supposed to bring out their data sets and say, “While we have not released them openly yet, these are the ones we have,” so that people can actually know what they have. I do not know if it has been updated now, but Treasury had a total of eight data sets that it knew of that had not been opened up. I would say that Treasury has been a blocker. The Treasury itself does not have data sets, but it has been a major blocker of much of this. That is reversed from the position before, where it was getting to the point where it understood the trade-off between giving up a small amount of jam today for a lot of jam tomorrow.

Q62 Chair: Could you give us in writing examples of the data sets you think they should be releasing as open data but that they are not?

Dr Pollock: To be clear, the Treasury does not have them, but it is often a deciding factor, because it is the one who has to say, “Okay, the postcode address file. That will maybe knock £10 million total off the amount we sell Royal Mail for, but it will generate £100 million a year for the UK economy, but we won’t do that because we’re so worried about jam today.”

Q63 Chair: Will you send us a note about that?

Dr Pollock: I will do my best, absolutely.

Q64 Chair: Presumably, there needs to be a regular audit of open data, as you have all indicated. It needs to be published so that we can see how well each public body is doing. Would you support such a recommendation? How would that be implemented?

Heather Savory: In its response to the Shakespeare review, the Cabinet Office has composed the National Information Infrastructure and is actively working on that at the moment. That will be made public at the end of this month at the Open Government Partnership Conference. That will go a long way to improving the tracking mechanisms. I work in association with the Cabinet Office. It is working very hard now with Departments to actually get this inventory out. That is very important because we put in place a data request mechanism that demonstrates user demand. That is what the Open Data User Group has focused its work on: what users need. That is the demand side, but this will articulate the supply side. I would say, though, that this will only be the first step. Anybody who says at the end of this month, “Here it is—the National Information Infrastructure is delivered,” is deluded, because there is more work to do to identify what data is there.

Q65 Chair: Presumably that will be an endless task.

Heather Savory: It will. Yes, it will be an endless task until the point that procurement rules include the presumption to get your data sorted out up front. In the medium to long term, I cannot see where the oversight is going to come from when this is embedded. At the moment, you have different people in different policy areas. At the end of the day, in two or three years' time, somebody needs to be given there responsibility to make sure that this embedded system, as it will be by then, is maintained for the long-term public good.

Q66 Chair: Could this be a role for UKSA and ONS?

Dr Pollock: On the regulatory side, this is one of the old ones. This was a recommendation in *Models of Public Sector Information Provision via Trading Funds*, which I, Professor Bently and Professor Newbery at Cambridge wrote. It has never been implemented, and the UK Government were instrumental in having it cut out of the revised PSI Directive, which was the idea of an independent regulator of a bunch of this. For example, on trading funds, a classic thing is that if you are going to allow some entity to provide information, you need to regulate quality. Ordnance Survey, for example, has to provide maps for the UK, and you are going to say, "Maybe we are either going to give you a subsidy now or we're going to provide some set of money." The Government already pay it £50 million a year.

The worry that trading funds have often had is if they are dependent on Government money, what happens in the bad times? In the bad times, everyone gets their hair cut, and suddenly you cannot do your mapping anymore. One way to run that is to have an independent regulator who both guarantees that they are at arm's length—they get provided money by the Government—but at the same time monitors quality and other things. There is a question at the moment of what happens if people complain? At the moment, they can come to the Open Data User Group and we can go to the Cabinet Office and complain, but it is not a very enforceable structure regarding access to some of these things. You have the Information Commissioner and things around other stuff, and one of the recommendations is having some proper regulatory oversight of this going forward that we do not really have today.

Tom Steinberg: This was the issue over which I decided that I would no longer be on the Transparency Board. A group of six civil servants from different Departments got on the telephone to tell me that the idea of independent regulation was not worthy for what are essentially entities—trading funds that, if they want, can just shave margin off every business and bit of government in the country with no challenge. They told me it was a stupid and bureaucratic idea.

Q67 Chair: Six civil servants sounds like a campaign.

Tom Steinberg: You asked me before which Department was most problematic, and I gave a non-answer, because I said Parliament, which is not a Department. Actually, I do think that was largely orchestrated by the Treasury. They were the ones who were most opposed to the idea of more regulation, because they were in the business of making less regulation. I understand that there are ideological issues about more or less regulation, but if I were running a country, I would not want entities that had the ability to suck the margins out of every business in the country to do so unregulated. In my view, that is the situation today. In the view of Whitehall, these entities are regulated—just by Whitehall Departments. If

that is the question, I would ask why we have any regulators at all. Why is every singly utility in the country not just directly regulated out of the Home Office or BIS and so on? The answer is because there is a different character and quality of analysis to independent regulators. These trading funds desperately need to be run by independent regulators. The fact that I could not move the Government at all on that was why I quit.

Chair: That comment has made a very strong impression on me.

Dr Pollock: Very briefly, *Models of Public Sector Information Provision via Trading Funds*, the 2008 report for the Treasury, has a whole chapter on regulation. It makes a very clear recommendation at the end. It is very simple and distils why you should have this. There are also the comments of the trading funds in readable prose.

Chair: Very good. We will look at that. Thank you very much.

Q68 Kelvin Hopkins: I have just a couple of comments on what Tom Steinberg has said particularly. I am most interested in what he last said, but his previous comment about the future we face with the direction of Governments moving towards a completely privatised world suggests we might get to a point when there is almost no way of judging how the world is operating. Democracy itself seems to be a bit of a nonsense because we cannot make any judgments on what is happening. That is a thought. Before that, you suggested that we did not need a central, standardised way of making sure that local authorities delivered their data in comparable or accessible ways. I just wondered if that was really what you were saying because it strikes me as being a very good idea, so that local authorities cannot hide their bad performance if they put their data out in a way that is not comparable with other authorities.

Tom Steinberg: Standards for data are important, especially if, as I said, the main lever for change is procurement. I really believe the main lever is that all new computer systems should be embedding these standards when they come off the shelf. My only difference of opinion there is that I do not think the British Government should decide that they have to own, run, maintain and invent all these standards. There are many things this Government do that many other Governments are doing and, therefore, just like we have comparable unemployment rates, we should in the future have data standards embedded in the software Government buys. When a piece of spending, a contract or performance data measuring the performance of some contract comes out, it should come out in the same format here, in Bolivia, New Zealand and everywhere else that kind of computer system is installed. There should be standards, but not organised only by the British Government.

Dr Pollock: There is just a quick thing I want to add. The danger Tom has not alluded to is that the Government will regulate and say, “You must use standard X,” but technology moves quickly. In three years’ time, that standard may be superseded, so there is a significant danger of embedding some precise rule into legislation in technologically rapidly moving areas. It would be better to say, “Use best practice,” whatever that is.

Q69 Kelvin Hopkins: Just one example of where there seems to be no agreement is on what we call the tax gap—tax avoidance; tax evasion. There are some estimates that put it at well over £100 billion; HMRC puts it much lower than that. There seems to be no acceptance of a common view on what the tax gap is. If there were some standardised way of measuring it or defining it, we would actually be better off in terms of information. Anyway, that is not my

question. My question really is: the review by Stephan Shakespeare of YouGov suggests that the Government should have twin-track policy for data release that recognises that the perfect should not be the enemy of the good. They should simultaneously publish early even if imperfect data, but at the same time with a commitment to what he calls a “high quality core”. Do you think that is sensible?

Tom Steinberg: GDP, which is probably the single number that people in Whitehall care about more than anything else, is published and then revised infinitely, yet there have been much less important data sets that have not been released because we cannot make a mistake. I would say to make one of two calls. Either do not publish GDP until you are really confident—in other words, about two and a half years after it is relevant—or be consistent in the Government’s approach to data and say, “The public can cope with the idea that it is not perfect, and the public can cope with the idea that it is going to iterate over time and get better.” I am very much in favour of the second position.

Dr Pollock: Sadly, it is almost always used as an excuse to not publish information: “It’s not perfect yet. People will misuse it. They will drive off cliffs because the geographic information has a road leading off a cliff.” I have heard that genuinely from someone before.

Q70 Kelvin Hopkins: There is the phrase “Governments wishing to bury bad news”—they can use that excuse by not publishing things that might embarrass them, for example.

Heather Savory: Can I add to that? The answer is to publish the result and publish the underlying data, and let other people look at your calculation or look at different ways of coming up with things to compare the result. The ONS does this. It now publishes statistics and there is a move, which is hugely welcomed, within the ONS to publish the underlying data sets that those statistics are derived from. On the questions of how we should measure or how we should calculate, you can pick any way, but just open up. This is where it is about transparency for government: open that up and let other clever people—because there are clever people all over the place—have a look at what you have done and understand maybe the limitations of your calculation or suggest a better way. This is why this is so powerful.

Q71 Kelvin Hopkins: The GDP revisions tend to be very small but, on the other hand, if it is the difference between zero growth and 0.1% growth, Governments will claim success.

Chair: There was no double-dip. We thought there was a double-dip and there was no double-dip.

Kelvin Hopkins: Indeed. On other things, there are massive differences. I raise these constantly; the tax gap is one of my obsessions, but it is almost never mentioned by comparison with the tiny amount we lose from benefit fraud, which is made much of. The tax gap, which is gigantic, is played down. It is politics. The one other question I had is how we lessen the risks posed to privacy by open data.

Dr Pollock: It is a really important issue. Talking earlier, I said something like we should make a principle of open data by default for non-personal data. I really reiterate this, so that by default when we talk about open data, we are obviously not talking about private

data. The risk is in what Heather said, for example. Much of the data underlying national statistics is information from individual people or organisations sometimes. One principle is that open data is not stuff that has privacy implications. We do also want to think quite carefully. Crime statistics are a good example. One of the interesting things is the variation. In the United States, they quite happily publish granular crime statistics of exactly where the crime happened, where it was and what time, but obviously not the exact person. Whereas in the UK, there was this whole debate that I could work out where this had occurred—had a rape happened in a house, I would know something very significant personally about someone. There is clearly going to be ongoing debate.

I do not think there is an easy answer to this, and I think it is going to be one where some of the most interesting data will have a relationship with personal information—crime statistics or crime data, in fact. The default position has to be that we protect privacy in the first instance, but it is important that there are cases where we make public interest trade-offs. We think that we are entitled to know the directors of public companies; it is not something that is private. By having a limited liability company, you give up your right to privacy—people know that you are a director. Those are going to be interesting questions. The really interesting question over the next 10, 20 or 30 years is how we make the trade-offs between privacy and public interest in a variety of areas. One has been around the VAT register—should the VAT register not be published because it possibly has privacy implications?

Q72 Kelvin Hopkins: There are some countries, such as in Scandinavia, where they publish people's personal tax returns and so on. That sort of privacy is thought to be unacceptable in a democracy—that we ought to know. We tend to be much more worried about privacy.

Chair: Can we let the other witnesses answer your original question?

Heather Savory: I agree with you that it is a cultural issue. I am quite interested to see what happens in the next generation, because they just work in such a different way. They communicate differently. They have a lot of private information available and seem to be far less fazed by it all, although that could turn around.

The VAT register is quite interesting because it is a good case study of what happens with Departments. We have this with the VAT register and we have it with DVLA data as well. The thing is, “We can't open up the data because we'll increase fraud, and personal data will be there.” This is coming back to the attitudes about it. With the VAT register, we found there is actually an EU site you can go to, VIES, where you can go and type any member state, VAT number, and it gives you the address. Sometimes you find that these issues of privacy have already been superseded. People do not actually know how much information is available about them. It is only when you start talking about releasing it as bulk data that people start to get concerned about it.

Having said that, we want open data but we do not want to compromise privacy. As Rufus has said, we have to make some decisions about how to anonymise data. There are some really quite serious working groups at the moment looking at anonymisation and

pseudonimisation of data so that you can release these data sets in a way that means they can be hugely beneficial, but without identifying individuals.

My last point is cultural. When you ask citizens to put their personal data up for a good cause like the Human Genome Project or various health projects, you find a totally different attitude. If people feel they are giving their data in a beneficial way, they are much more open to allowing it to be used for research, etc.

Q73 Chair: Can I just probe on one particular question, which is this generational difference? How do you think that can be reflected in Government policy, or does Government policy have to evolve with changing generational attitudes towards privacy?

Tom Steinberg: Yes. Step one is accepting the fact that values in this field are changing really fast and changing in ways we do not fully understand yet. That means understanding that, unlike legislation for a crime, where you are pretty sure people will think that is wrong today and still think it wrong in 20 years' time, instead you have to think of this much more like something that is inevitably going to change quite quickly. I would argue that if Parliament has a role here, it is probably to make sure that there is some kind of regular review that is much faster than normal, decade-long legislative cycles and can cope with the fact that what is acceptable today may even have changed in acceptability in just two or three years' time. I would like to see Parliament say every 12 months or every 24 months the Government are essentially going to be tasked to assess what is normal, or Parliament itself will meet to discuss what is now acceptable and what is not acceptable in this trade-off between privacy and public. The only thing that is permanent is the decision to review it very regularly in a way that the rest of the public sector can understand.

The public sector really needs some guidance that lands on their desks that basically goes, "These kinds of trade-offs are okay and these kinds are not," because you are never going to be able to draw a crystal-clear legal definitional line between them. It is not going to exist in this field; it is all shades of grey. What you can give people, though, is five examples of what Members here vote in a majority to say are acceptable trade-offs and five that are not. Review those in two years and there might be a different five and a different five. I would love to see a decade in which we prepared for the knowledge that these values are going to change so quickly and that the political and technical context is going to change very quickly, and that in the diary we know when it is going to be next revised, even though it has not happened yet. That would be a great step forward.

Q74 Mr Reed: I will just confine myself to one question.

Chair: That is economic innovation and so on. Mr Pollock, can you take the first answer because I know you have to leave?

Mr Reed: How well do the Government understand the needs of different commercial users of open data?

Dr Pollock: It is a difficult question because there are so many. Heather has emphasised that sometimes it is quite well known what some of those desires are, but the challenge is often they are the ones the Government are least inclined to satisfy for other reasons. I

think the Government do have a sense of that, but a key point is that much of this is about innovation. It is always a challenge of doing stuff: maybe it is people who are 10 or 15 today who are going to build a company in two, three or five years. It is very difficult because they are not normally the people who come to this room or even know yet what their wants are.

One of our lessons of innovation, particularly in the digital space, is that you do not know what you are going to build, and often when you think you know what you are going to build, you are wrong. In that sense, the whole point of opening up data is that it is the unexpected. It is the many-minds principle. The best thing to do with your data will be thought of by someone else, not by you. That aspect is essential to this. In a way, it is currently a debate about high-value/low-value data sets. There are some things we can see now there are desires for, but some of the things that are most interesting we are not going to anticipate. Maybe someone will come up with a credible way to spot tax evasion or tax avoidance, or better ways to look at benefit fraud, or better ways to identify how government saves money. Those are the most unexpected aspects of this, and we are not sure where they will come from. It may be some other data set that we cannot even think of.

That is one of the fundamental points here, in the sense that it is very difficult for commercial entities. Google could come and tell us, but those that are most exciting and most interesting I think would find it very difficult to come and tell us. This is why “get it out there” is the general approach.

Q75 Mr Reed: Is there any way that government can engage with that kind of dynamic, entrepreneurial fast-moving world?

Dr Pollock: Go to events. Get out of government and do not run consultations. That is my other big point. People who write responses to consultations, even me, are busy—technical innovators are busy. Go to where they are. Go to their meet-ups, go to their events, go to their conferences and listen to them. Unfortunately, sometimes the Government in other times have said, “No one has a travel budget to get anywhere.” Go to those things; it sometimes means people have to do it in their spare time. They happen in evenings or weekends, or require you to travel to another country. Go to where they are and listen to them.

Heather Savory: That is actually one of the reasons I am here, because obviously the Open Data User Group was set up to try to improve that. In terms of Government understanding, I think one of the things the Government do not understand is how big companies need this data. I have something to wave at you as an example. This is done by Deloitte, and it is using open data. It is a demographic analysis of how population is moving and where current bank branch networks are. Nobody likes the banks much at one level, but they are very important in our society. This shows that retired people are all moving to the seaside, and it shows that young, upwardly mobile people are moving into cities. This is based on open data. This is why companies that you do not think of as needing open data need this open data. These are the utilities companies, Sainsbury’s, Tesco, etc.

On the start-up side, we need to make more of case studies. One of the things I want to do this year is to try to promote some of the really good examples of businesses built on

open data. We need to gain confidence, particularly within the Treasury, so that this out-of-balance “small amount of money for big benefit” could be tipped when they start to see some tangible results.

The primary thing, which comes back to what both my colleagues here have said, is about pace. Government just does not understand the difference in pace between Whitehall and Departments, and then somewhere along the line there are other public sector bodies, and the business world, and businesses of all size. For example, we have had 500 data requests since last year. They are all moving slowly through the system and some of them are now coming to fruition. What these start-up companies want is if it is a “no” because of private data issues, it needs to be “no” tomorrow, because then they will go and do something different. Then in Whitehall everyone is saying, “You’re not doing badly, are you? Things are moving really quickly, aren’t they?” and then I have my group saying, “This is glacial.” There are two cultures there.

Q76 Chair: Do you think this is all tainted by a Sir Humphrey disdain for the idea that people should make money out of public data—that there is something rather distasteful about it?

Heather Savory: There are definitely some individuals I have met during the course of the last year who do not seem to understand the fundamental connection between making money, paying tax and delivering public services. I found that quite shocking. This is a virtuous circle.

Tom Steinberg: Can I just answer your question? I slightly take issue with the idea that the Government should need to understand much about me. If I have truly amazing, wonderful new idea, it is entirely likely that no one understands it except me, let alone the people on the other side of the table. Luckily they got rid of this, but it used to be that the Ordnance Survey asked for your business model so that it could read it and understand your need. At one level, that sounds like it makes sense, but at another level the moment you say the Government need to understand what someone needs some data for is the moment you say, “Our Government’s data policy has basically excluded true innovation.”

The nature of innovation is often to see value where other people just cannot see it. No one saw that there was much value in a list of Harvard undergraduates except Mark Zuckerberg; no one saw there was much value in the links on the internet pointing to other websites except the people who founded Google. A characteristic of some digital innovation is precisely people seeing value. If the Government are set up to say, “We’ll give it to you if we understand the value,” that is just basically a way of saying, “We wouldn’t like any innovative companies, please.” I would like it to be thought of in a rights model much more than in an understanding model. That is to say: if I tell you that it is important to me and it does not place an unreasonable burden on the public sector, I will get it. Merely by coincidence, that is how the Freedom of Information Act is structured at the moment. It does not work, though, if I go and ask the Government for some bit of data. That is very much how it needs to work, precisely with the 18-year-old Rufus talked about, who is not going to know how to make a value case to a civil servant in civil servant language.

Dr Pollock: You need to connect to the electricity grid: “Can you tell us why you need electricity to run your factory?”

Q77 Chair: You would be surprised.

Can I ask one final question? Stephan Shakespeare talks about how our country could be at the front of a wave of innovation that would provide another big service industry we would export all over the world as other countries move to open data. Do you agree with that analysis or that conjecture? Could we be the home of all the new apps, applications and development?

Dr Pollock: It is definitely possible. There are big forces at work, but one example often given of where we failed was that the US opened up geographic information in the '80s. It was very liberal about access to geo-information and we were not; we actually shut it down under Thatcher by charging more and Ordnance Survey moving to the trading fund model quite aggressively. It is always cited; there is the famous saying by Weiss where he is saying in the US you have this really big geographic information system industry. There are people selling geographic software, not selling data. There is also the example of weather derivatives as well in the US, where access to free weather data made it much easier to build a weather derivatives market, where you could insure weather risk much better and more efficiently. There is that possibility. The world moves fast, so we need to do that. There is potential. There is going to be competition in that space, and it relates to the software industry and other things. It relates to having a good venture capital industry and having a good set of people who are trained in data and information skills. We run the School of Data because we think all this data coming out is great, but do people know how to use it, particularly people who are not data geeks? That is the other challenge. It is not just going to be the data out there; it is going to be the skills, the infrastructure and the funding to build the businesses to do that as well.

Chair: I think that was a yes.

Dr Pollock: Yes, that was a yes.

Heather Savory: I got slightly confused by your question because I think there are two aspects to this. There is the UK's skill at open data, which is what I suddenly thought you might be asking. To what extent can we influence other countries and share our best practice in terms of what we do here? I think the answer there is yes, because we are at the front of the field. I agree with what Rufus is saying there. There is absolutely a skill shortage. A lot of techy friends are now delighted because somebody has said that being a data scientist is going to be the sexiest job in the 21st century. That is actually true: there are opportunities. We need those skills. We also need to make sure that the financiers really can depend on this source of material. That is fundamental. I am not going to back a business if I do not think that this data is going to be available in two years' time.

Q78 Chair: The rights-based model rather than a permission-based model. I understand that.

Tom Steinberg: I think we have almost everything we need to be building the 2013 versions of the 1990s and '80s tools that were mentioned there. You mentioned earlier on the idea of a tool that could tell you how much public money had been spent in an area.

Someone can get paid to build that tool and someone can get paid for selling it to Governments around the world. It will get built in the country where the data to power it is available, and right now that is not Great Britain. The reason it is not Great Britain is because the Treasury does not believe that is a way that we will generate economic growth. It just fundamentally does not buy the argument at a very deep level. One of the things that I have found mostly impossible is to change its mind on that issue, and that is why I am no longer actively involved in this field.

Q79 Mr Reed: Is it worse than you have just outlined, in that the argument is that it sees negative outcomes from it?

Tom Steinberg: It sees a loss of short-term revenue, yes.

Heather Savory: There is a negative outcome, because they will lose certain income streams. That is known. That is why the cost-benefit argument is there. It is what I said earlier: it is the funding model versus the economic model. We also have to have faith in the UK. We have done some great things, and this is a fundamentally easy place for the UK to generate growth. That is what we need people to see. You do not need huge capital infrastructure. I have a guy who took the INSPIRE database. It is a huge database. The computing power that you can get for £10,000 will allow you to do an awful lot of development with this open data. The difficulties that you have are finding the skills and paying the people to do it. That is the side of this that needs to be addressed.

I cannot see many other places. There are other places where we are strong in sciences, etc., but if you look at the potential for growth that we have here, we do have a platform to build on in the UK. We have great people; we have great technologists and great scientists; we have great brains. This does not require you to build a factory. I came from the semiconductor industry; we used to have semiconductor fabs here and they all moved to the Far East. They moved to the Far East not just because of the costing, but because as technology improved, you needed purer water—there are geographical factors there as well. There are none of those here. It is about people and computers. We have those.

Chair: It has been a really good evidence session. You are all fantastic witnesses. Thank you very much indeed, and if you see our subsequent evidence session and have any further thoughts, please do whack in any further comments that you might like. We will certainly take heed of them. Thank you very much indeed.