DATA SHARING BY GOVERNMENT
ABRIDGED VERSION OF KEYNOTE PRESENTATION TO DEMOGRAPHICS
USERS GROUP ANNUAL CONFERENCE
ROYAL SOCIETY
2 OCTOBER 2008

I have been given a very wide ranging topic for this presentation. So, at risk of disappointment, I first want to narrow it down a bit. I am not going to be talking about inadvertent data sharing by government. Discs in the post, laptops on the train, missing memory sticks. That sort of thing. If you are interested in that I can commend the recent report of the Joint Committee on Human Rights on Data Protection and Human Rights [HC132].

I am also not going to talk about exactly who decides what gets captured on CCTV, who is on the DNA database or who can look at our medical records on the new NHS computer. Those with an interest can take their pick of any number of conspiracy theory websites.

I am going to focus on the issue of data sharing by government as it affects the topic of your conference today. Data sharing to get insights into our economy and society.

I was helped in my preparation for this presentation by the recent report from Richard Thomas, the Information Commissioner, and Mark Walport, the boss of the Wellcome Trust. They were commissioned last autumn by the Prime Minister to conduct a review of data sharing. They devoted a chapter of their report to making some critical distinctions that are highly relevant to our discussion.

They categorised data sharing into three groups. First there is data sharing for the purpose of law enforcement or public protection. This deals with the use by the state of information about individuals where there is a potential conflict between the interests of the individual and society at large. Difficult issues arise and reasonable people can hold different views. For example, is it right for the government to give utility companies access to information about the income of individuals without their consent in order to help alleviate fuel poverty? Or another example, who must a doctor tell if he believes a patient is likely to be dangerous to others? In a democratic society it is the job of Parliament to debate, decide and keep under review each instance when a choice like this has to be made.

The second category is data sharing for the purposes of service provision. This deals with the use by an individual of information relevant to that individual but held by others. My favourite example is the website for renewing car tax. This makes my life easier. I can log on and organise sharing of data about my car, my insurance cover and my car tax. I decide who gets to see what and can elect to do the transaction over the counter in the post office if I don't like the website method. I am in control of who gets to see what. For my children this approach operates on a much larger scale in sites like FaceBook. I sense that there may be a generational change in the willingness of individuals to share information in this way. Time will tell, but the crucial element of this kind of data sharing is that it is controlled by the individual.

The third category, and the one that is important to us today, is data sharing for the purposes of research and statistics. This deals with the use of aggregated and anonymised information about individuals. Individual data needs to be processed but only for the purpose of creating new information which does not relate directly to the individual and does not allow others to infer any information about any particular individual. Is there an increased risk of leukaemia amongst those who live close to high voltage power lines? Which factors are linked amongst households facing multiple disadvantage? How many people live in a particular place and what are their demographic characteristics? Answers to all of these questions require this kind of data sharing.

These distinctions are critically important but are not as widely understood as they need to be. Indeed there is a serious risk that essential statistical analysis is prevented by an environment that will only share data for statistical purposes if the same kinds of legal safeguards are in place as are needed for the purposes of law enforcement and public protection. At the same time there is a serious risk that the same statistical analysis is prevented by an environment that will only share data for statistical analysis if each individual gives consent for each individual use.

There are already examples where this kind of risk has been observed. Analysis of cancer survival depends on the sharing of data about individuals and it is not always practical to get consent. Unless analysts have confidence that samples represent the true populations the validity of results can be compromised. I believe that very few people would want data to be managed in a way that prevents good analysis of cancer survival yet an environment that insists on consent for all uses could have this result.

Another pertinent example is the scope to develop the government's administrative databases to provide a more accurate and cost effective method of counting the population, as recommended by the recent House of Commons Treasury Committee Report on Counting the Population [HC 183].

The Data Protection Act 1998 recognises the special place of research and statistics and the Statistics and Registration Services Act 2007 provides an opportunity to take this further, but the job of everyone in this room is to make the case and demonstrate that the benefits of doing so outweigh the risks.

The task, however, is not an easy one. In the last 20 years the landscape of data sharing in statistics has, I think, undergone two dramatic transformations. I will illustrate this by telling stories from three general elections ten years apart: 1987, 1997 and 2007. Now 2007 was, of course, the election that wasn't, but the fact that it wasn't was, in some people's estimation, due at least in part to issues of data sharing. In 1987 we inhabited an agricultural economy for data. By 1997 we had become industrialised. By 2007 the service sector was dominant. Let me explain.

My clearest recollection of the 1987 election was Margaret Thatcher and her handbag on a grotty bit of land in Middlesbrough announcing to the nation that we needed to get back into those inner cities. This mattered to me because I was working on inner cities policy in the Department of the Environment.

In implementing the policy laid out in the election campaign our data challenge was quite straightforward. We were hunter-gatherers, seeking information that was already around, or were farmers, cultivating our own statistics through specially designed surveys. We may also have been traders, exchanging what we had with our friends and colleagues.

In the development of London Docklands I can recall a helicopter flight hovering where the top of Canary Wharf was to be. Success was easy to judge – a tape measure to see whether the Reichmann brothers' edifice was as high a promised, some counts of the square feet of office space, the number of jobs, the traffic on the Docklands Light Railway.

I can recall a trip to the Black Country pausing at a housing estate overlooking a lake. As the sun went down and the light faded the light of the water did not fade. The lake glowed. It was on the site of an old match factory and was full of phosphorus or somesuch. We could judge success by how well it was cleaned up. Less glowing. Some measurements from the hydrographers. In hindsight, 1987 was an agricultural idyll for data.

Jump forward to 1997 and we have Tony Blair on the campaign trail telling us that the economy was prospering but that some in our society were getting left behind. Something needed to be done. This mattered to me because after the election I was closely involved with the Social Exclusion Unit set up by No 10 to do something. The first thing they wanted to know was what is this social exclusion and where can it be found. Answer came there none.

In between 1987 and 1997 data had become industrialised. Computers were now commonplace and great data warehouses were in place processing vast arrays of administrative information. The Social Exclusion Unit set up three teams to explore how best to get hold of the data to answer its questions. It is great that the three people who chaired those groups are here today. Bob Barr had the job of looking at geography. I had the job of exploring issues of data sharing and privacy. And Keith Dugmore led on identifying the data that could be used.

Our conclusion was that, generally speaking, the data existed and could lawfully be accessed for these purposes. The impediments were primarily organisational. We just needed to get someone to set up a processing plant which could organise the supply chains, add value to the data and repackage it for use in various communities up and down the country. ONS was given this task and the result was neighbourhood statistics. We used the prevailing industrial model to bring together the available data and manufacture something new and useful out of it. In hindsight, 1997 was such a straightforward world.

Onwards again to 2007. So much happened in the intervening 10 years. For me the moment of the change from an industrial to a service orientated world for data was 29 April 2001. 29 April was the day of the 2001 census and I had my work cut out explaining why people should trust the census authorities with their data. The problem became much greater again later that year following September 11 as governments sought to explore all methods to obtain the information needed to combat the threats to national security. Each year since, the issues have compounded. Data farmers are

still at work in the fields and we rely on them. Data manufacturers are still processing in their factories and we rely on them too. But the big debate has been about who we trust. Which information "brands" can be trusted to collect information from us and keep it safe, to process it in ways we feel comfortable with and to serve it up to us in forms that we believe are straight and unspun.

So by 2007 we are in a service economy where value is created by those brands which effectively differentiate themselves in a crowded market and stand for the things that people care about. The Statistics and Registration Services Act 2007 is about that process of differentiation. About creating a brand that builds trust. The Thomas/Walport review stressed more generally the importance of creating "safe havens" for data and the idea of "accreditation" for those who can be trusted to do certain things with it.

The task is not an easy one. Whilst, or perhaps because, the UK is arguably one of the most data rich countries in the world, the public are unconvinced about whether those holding their data can be trusted. Thomas and Walport cite a Eurobarometer survey which showed that 77 per cent of the British public are concerned that organisations cannot be trusted with their data, much higher than the average amongst EU states.

If we need to bring together information through the sharing of government databases on order to provide important insights into the functioning of our economy and society we need to make the case. This new world which we inhabit is data rich but trust poor. All of us in this room have an interest in tackling this issue head on to make sure that our richness in data is a blessing not a curse.

John Pullinger