

Speech by Stephen Penneck, Director General, Office of National Statistics to The Foundation for Science and Technology on 8th June, 2011

Introduction

The question tonight is “Can better use be made of public data?” Although the framing of the question might instantly draw out a simple yes, the journey to the answer is not in fact a simple one. Whilst it may be hard to disagree with the question in principle, in practice it is equally difficult to offer a simple ‘yes’ without highlighting a few of the public data hazards on the road to greater use and wider access.

This evening I am going to look at the issue from the point of view of ONS. First I will set the scene before moving on to:

- outline some of the challenges and opportunities we face; then I will
- explain how the Office for National Statistics is responding now; and finally I will
- look ahead to how we are aligning our future strategy with this new agenda.

ONS has a vital role to play in both promoting our statistics as public data and protecting their provenance. I think National Statistics differ from other statistics. Firstly in their importance: National Statistics shape all our lives. Secondly, the metadata we publish enables users to understand their origins. Thirdly, the National Statistics brand enables users to trust their provenance. The brand allows users to pinpoint statistics they can trust and distinguish them from all the other data they can find on the internet. From inflation to migration, National Statistics gathered from public sources illuminate today’s information superhighway. They show us the route society is taking, picking out both the direction of travel and highlighting the occasional pothole. In effect, they light up the past so that we might glimpse the road ahead – a road where public data is underpinned by public trust.

At the Office for National Statistics, making public data accessible is at the heart of everything we do. As the UK’s recognised national statistics institute, ONS embraces the transparency agenda. It is in the public’s interest that we support the ‘right’ to public data. But ONS data without provenance makes transparency opaque.

We are the UK’s largest producer of national statistics. Last year, for example, we published nearly 300 statistical bulletins and 100 news releases. But we don’t just publish our numbers, we explain them too. We host media briefings, and organise an array of talks and specialist seminars on a range of themes, ranging from market-sensitive economic and business data to information on health and societal well-being. In addition, we play a leading role in the development of national and international best practice, maintaining our position as a world-class provider of statistics. All of this is important if users are to make the best use of our statistics.

Numbers may show us the past, but we look to the future. There has been a major shift in the way ONS works with users in recent years, away from narrow paternalism towards the broader idea of what market researchers might call the 'citizen-consumer'. Once we acted as custodians, protecting public data. We decided which datasets might be of public value, and published them in support of our official statistical releases. Now, we live in a world where our customers can ask for the publication of datasets which do not reveal individual and personal information. Putting the public at the centre of our statistical services fully supports the transparency agenda.

Background

Twenty years ago, few people would have predicted how widely statistics would illuminate our lives, or how widely public data would be published today. 1991 was marked by major world events, including the collapse of the Soviet Union and the first Gulf War. But one historic happening quietly reached further into the collective global consciousness than any other. According to the UN, it went on to affect the lives of 1.6 billion people – almost a quarter of the world's population. Yes, on the sixth of August 1991, Sir Tim Berners-Lee put the first ever website on the internet.

The following two decades have seen the internet become a phenomenon; changing the way we work, learn, communicate, shop, make friends and even meet partners. This year, online dating agencies made it into our Consumer Price Index 'Basket of Goods' for the very first time, along with 'smart phones' and their 'apps.' It seems funny how the fax machine and floppy disk, once the mainstay of any modern business, now seem like figments of our imagination ... that is, of course, if you're over 40.

During a speech last year on 'The Next Age of Government', the Prime Minister recognised that the internet will continue to change how government works, improving transparency, increasing choice and encouraging greater accountability. Recently, the Government's digital champion, Martha Lane Fox, sought to engage millions more Britons online with the launch of the Race Online 2012 campaign.

The scope for publishing public data on the internet is vast. Evidence gathered by ONS in a recent Social Trends report on e-Society showed that, in 2010, nearly three-quarters of households in the UK had internet access. Our figures show that thirty million UK citizens go online every day – or almost every day. Almost half of all women who go online, and 40 per cent of men, use a social networking profile. Ten per cent of us work from home using the internet. And more than sixty per cent of our children do their homework online at least once a week.

Challenges for ONS in Making Public Data Public

So what are the challenges that the Office for National Statistics faces in making public data more accessible? Sir Tim Berners-Lee said last year that making public data available is about increasing accountability and transparency, and letting people use data in new, innovative ways. The Government believes improving access to public data will help citizens make informed choices about public services, and make them better able to converse and collaborate with Government.

The first challenge is to have a clear definition of public data. The Transparency Board chaired by Francis Maude has defined “public data” as “non-personal data on which public services are run and assessed and on which policy decisions are based”. Clearly this covers ONS data.

So what are the challenges?

Data Security and Access

It is important to stress that the agenda here is about non-personal data. We go to great lengths to safeguard the information provided by individuals and businesses – it is as important to us as it is to those individuals and businesses. Protecting the confidentiality of the information in our care is a vital part of all our business practices. We do not publish anything that is likely to identify an individual, household or business. We include a confidentiality guarantee on all of our survey questionnaires and on our survey literature. And most importantly, we adhere to the Code of Practice for Official Statistics.

But this does not mean that such data cannot be exploited more fully as long as we also safeguard confidentiality and earn public trust in everything we do. So it is not just our commitment and ability to make public data public, but it is also about our commitment to keep private data private.

Census 2011

Nowhere is the duality of these obligations more pronounced than the Census – arguably THE public data event of the decade. The 2011 Census questionnaire went out to 26 million households across England and Wales – and well over 90 per cent of them have now completed and returned their questionnaires.

Throughout the whole of the Census process, data security is a top priority. Indeed, ONS confirms its overriding commitment to ensuring the confidentiality of personal Census data for a period of 100 years.

Not only that. All employees of ONS, and all contractors working with Census data, are bound by the Census Act, which dates back more than 90 years. In essence, any breach of confidentiality is a criminal offence, subject to possible imprisonment and fines. Census staff sign a confidentiality declaration to make sure they have understood their commitment to this form of legal sanction.

The Census is by far our biggest snapshot of life in the UK, but we conduct many others. Our business surveys, for instance, collect information that helps measure the state of the wider economy, and helps the Bank of England set interest rates. And the Annual Survey of Hours and Earnings is used to make recommendations about the minimum wage.

Having said much on the confidentiality of personal data, let me reassure you that we do want to make the most of the information we hold.

We know that unpublished data can be a valuable resource for academic researchers and for some time we have been making our data available to the research community via a number of routes.

Let me say more about what we have recently been doing. For ONS the drive to make our data more accessible has been influencing our plans for some years.

VML / SDS / data archive

For example, we provide access to approved researchers under strictly controlled conditions via our Virtual Microdata Laboratory or “VML” based on ONS sites. The VML has been described by the OECD as “one of the most innovative research efforts in the public sector across the 30 OECD member countries”.

The ESRC has recently launched the Secure Data Service which enables safe and secure remote access by researchers to ONS data previously only available in the VML. The VML and the Secure Data Service are relatively recent developments compared with the UK Data Archive which was established over 40 years ago. ONS data is available from the Archive for the purposes of research and teaching, again with some restrictions.

However, not all users need such detailed datasets nor can they all meet our restrictive conditions. We have other solutions to help these users.

NOMIS and NeSS

ONS has been able to free up more than two billion statistics for public use through our Neighbourhood Statistics Service and NOMIS websites. In our consultation on our work programme, users said how much they valued these websites and facilities.

We publish much information on our own website too, including data in over 300 statistical bulletins, as well as time series data for the individual component series in downloadable form. But we acknowledge we need to do more.

From print to internet publications

However, before moving on to talk about the “more” we plan to do, let me say a little about our print publications and use of social media. Last year ONS moved all its printed materials to the internet, reflecting the need to broaden the use of official statistics and balance the decline in demand for weighty printed publications. By harnessing the power of the internet, ONS has been able to make data available to users more easily, whilst simultaneously reducing our printing costs and offering better value for money.

And of course we now “tweet” all our major releases. We have a Facebook page, and you can see some of our analysis on YouTube. From mobile ‘apps’ to software ‘data mash-ups,’ demand for the streaming of relevant and timely statistics is growing.

Our plans for the future

Outputs

So I have told you a little of how the public data agenda impacts on ONS, what the challenges are, and what we have done so far. Let me now tell you something of our plans for the next two years.

Our new website is being launched in August. It will include a data repository for published aggregate statistics, which, together with the planned Data Explorer and new APIs (Application Programming Interfaces), will support recognised open standard formats. Our new web facilities will also better support the dissemination of datasets and tabular and graphical outputs and we are certainly planning to put these to great use, by publishing more data sets. You may like to note that we are aiming to release the first census results (Local Authority population estimates) in July 2012.

While mentioning the Census, I would like to spend a couple of minutes sharing with you work we are carrying out on “Our Beyond 2011: UK-wide Population Statistics Programme” which could transform the way we currently work. This programme has been set up to establish and test models for meeting future user needs for census type statistics, and to address the implications of a transformed population statistics system for wider socio-demographic statistics. A particular aim of the programme is to test the enhanced use of administrative and survey data in future models, as an alternative to the UK's traditional census approach. Administrative data will improve the richness of the data available, especially for small areas, which must be a benefit to users.

As we take this work forward, maintaining public trust in data security standards will be absolutely critical as will the need to ensure public confidence in data suppliers and their data delivery.

Public scepticism that comes from combining and aggregating quantities of personal data across government, and the concern that these data may be misused, must be replaced with a trust in our ability to maintain the confidentiality and security of personal data.

There will also be a challenge for our statistical modelling techniques, and to explain to our users the quality of the resulting estimates. But this is some time in the future. In the meantime, we want to push forward our agenda to publish our public data so that it can be used and re-used by our users.

How are we going to do this? Our dissemination strategy for the future is built on a number of planks, recognising the varied uses of our statistics and analysis.

First plank

The first plank recognises the importance of citizens who want to take information from us in a form they can directly use in their daily lives. Statistical literacy is an important part of making better use of public data, to give people confidence to make sense of the data, or at least ask the right questions about it. That's why we

are working with the RSS on their GetStats campaign. To widen our user base among informed citizens we will simplify and improve the presentation of our statistical releases and improve our commentary. For example, we will publish more of the policy context for each of our releases and improve our graphical standards.

We will focus our analysis on where we can add value; for example:

- where we are in a good position to draw on a range of statistics to produce cross cutting analysis;
- where the analysis is dependent on a good knowledge of the quality of the statistics; and
- where we can provide helpful analysis at the time the statistics are published.

Another important aspect of making better use of public data is to use innovative methods of analysing and presenting existing data, using the power of the internet. ONS has pioneered such data visualisation methods in areas such as population pyramids and the personal inflation calculator. Looking forward, we intend to deliver the Census with some very powerful new methods of data visualisation.

Developing social media

Recognising the value of social media in reaching new users, we will continue to use and develop the social media channels we have set up over the past six months.

The use of mobile devices to access online content is increasing and likely to overtake access to online content from traditional desktops in the next three years. Making ONS content available through Facebook, YouTube and Twitter means that this content is available to users of mobile devices through these channels. But we will also look at the potential for providing a mobile stylesheet for the ONS website, so that users can browse these directly on their mobile devices.

Users of smartphone devices are increasingly choosing to access content through an "app". So we also want to investigate the potential for creating ONS mobile applications, which brings me to ...

Second plank - Third parties

Our second plank recognises that third parties are better placed to undertake some analysis and provide some services, and that we have a duty to supply the necessary data. So, we will publish on our new website in open data formats, enabling users to lift figures easily from our tables for further analysis. As I mentioned earlier, our new website will have facilities so that, over time, we can make low level aggregated data sets (all non-disclosive of course) available over the web, and technical users will be able to extract them for their own use through the API.

We have a good long-standing relationship with local authorities, health authorities, universities, charities, and professional bodies. In future, we want to collaborate with a wider range of partners, including the commercial sector, helping them create new products and services. After all, we have a wide range of data to offer for innovative use.

Third plank - Maintaining the VML / SDS

Our final plank is access for researchers to disclosive data. We plan to develop the service provided by the Secure Data Service, restricting the VML to data sets for which the SDS is not appropriate. We will continue to provide an effective service to enable Approved Researchers and other authorised users to access datasets where the appropriate permissions are in place.

Conclusion

In summary, then, ONS is striving to make public data public. We are developing our IT systems, building a repository for published statistics, improving our website, and opening up sources, in a spirit of transparency. But it will take time.

So how can we maintain public trust in this new world? If we are to transform the way we use public data across government and society, the road to change must come with appropriate signposts that reassure the public. We must act in their best interests, responding to public rights and safeguarding private data. Generations of statisticians and researchers can testify that public confidence is hard to win, and easy to lose. In short, whatever we make of public data, we must maintain confidentiality and protect its provenance. In the push for greater transparency of public data, accessibility is essential, but so is trust.