Workshop B: Data sharing – a public / private 2-way street?

We began the workshop with some introductory comments illustrating the status of each direction in the public/private two way street.

For example many organisations - including DUG members - find ONS's Census-type information very valuable. Most turn to commercial companies as suppliers of information. Currently 6 private sector businesses are assessed by ONS as Approved Suppliers of census data and value added information derived from the census. There are many more organisations that are Licenced Distributors of the data.

A decade ago the marketing data derived from the census was taught in university. Today teachers ask the census agencies to allow their GCSE pupils to access the data as part of business studies projects.

Such keen interest demonstrates that there are instances where public data already provides great value to businesses, the economy, and local communities. Extending the access to more public data can be expected to generate further interest and benefits.

What about data from the private sector to the public sector? The conference papers indicated how government is keen for ONS to investigate commercial sources of information that might be incorporated into public data.

Back in January ONS ran a workshop on Information Collected by Commercial Companies that might potentially be used to help replace the census. Around 20 commercial organisations attended.

What came out of that workshop?

Firstly the range of information collected about customers topics is very wide, extending well beyond those traditionally asked in censuses. However, coverage of the population is incomplete, particularly amongst some hard to reach groups.

Pooling of information already occurs in the private sector, sometimes with results being provided to the public sector. This may be driven compulsorily by regulation, or voluntarily for marketing purposes.

Companies were willing to help ONS where they could. There was a desire to minimise risk, effort & cost and any sharing of individual data is likely be prevented by Data Protection regulations and involve reputational risk.

Pete Benton expanded upon ONS' Beyond 2011 initiative, the purpose of which he summarised as "how best to provide information in the future".

There are three ways in which the private sector can assist ONS in this regard.

The first is helping to quantifying the benefits of public data. the business case for funding for the 2011 census was mainly based around the need for accurate population figures in order to effectively allocate resources, for example in the NHS and local government. Collecting the additional, more specific information, for example family structure, housing, occupation et al cost relatively little on top of the core process of obtaining the population counts.

Since Beyond 2011 seeks alternative and less expensive ways of obtaining the population counts it may well be that the additional detail will require more cost-justification in its own

right. The private sector often identify these variables to be of interest and the Beyond 2011 team would welcome (in confidence) case studies to assist them in evaluating the business case for demographic detail for small geographic areas.

The second type of assistance involves methodology. ONS are well-versed in their traditional skills of counting populations, carrying out surveys and processing large comprehensive data sets. The private sector has more expertise in the effective use of relatively sparse 'dirty' data, and ONS anticipate they may well seek advice and assistance from organisations such as those attending the conference.

The third area of assistance involves data. ONS are well aware of how difficult it would be for private sector organisations to pass data about their customers to government. Peter spoke of the possibility of aggregate, anonymised data, processing that might be done within private sector organisations rather than by ONS and other possibilities of creating information that would be of general benefit to the economy while ensuring businesses were protecting their customers' information.

Paul Munro from Barclays illustrated the data sharing that exists within the private sector, pointing out that in most cases the trick is sorting out the legal, compliance and administrative processes well in advance, the technical details tending to be well understood.

The first example was the various market databases in the financial sector. These involve many and sometimes most of the organisations in an industry sector pooling data, which is processed by a third party. For example Barclays and others would pool anonymised information for their current accounts. This would be processed to provide statistics about the 'total market' and each contributing member would receive data on their own accounts and the market as a whole.

Consumers give permission for retailers and others to pass their personal details to credit reference agencies. There are three main agencies and many organisations will use a mix of two of these. The transfer of personal data is within a 'secure controlled environment'. Since 2006 in addition to providing 'black' data such as CCJ's poor repayment history etc organisations now also pass over 'white' data, the number and amount of existing loans, the fact repayments are being regularly made.

Paul placed some value on this data sharing – "It's difficult to understand how organisations would work without it."

Nicky Tarry from the DWP then discussed data sharing within the public sector and the degree to which administrative data sets are already used to publish statistics at quite detailed level of geography.

It transpired that few attendees at the workshop were aware of the Output Area level data published on the DWP web site, or the utilities to customise your own data tabulations from the databases held by the DWP. In addition to counts of benefit claimants information was derived from National Insurance returns from employers. There is data in the flows on and off various benefits childcare, carers, pensions, and so on.

Nicky noted that the DWP run workshops on the availability of and requirements for data that are well attended by the public sector but rarely attract private sector interest. However there are instances of private sector data being merged into the public sector databases to good effect and there are instances of individual level data from HMRC being appended to the data.

The DWP would appreciate input on the statistics they publish, specifically identifying information that is of little interest since these impact of the cross tabulation with other variables and the result in limitations to ensure personal data cannot be deduced. Removing less desired data allows other variables to be published.

Following these introductory comments the workshop split into groups for wider discussion around private-public data sharing. Some of the points raised were:

What is in it for business? It was difficult to see extensive benefits for business. There appeared to be no significant income to be gained and while there was some positive PR in terms of socially responsible actions for the greater good these were felt to be outweighed by the risks of reputational damage in media portrayal sharing of data – even if the data were anonymised and avoided data protection issues. Moreover the data that might be shared has either been bought in at some expense or collected at some expense and sharing might in effect mean making it available to competitors for free.

Sharing data was seen to raise many issues including legal, data protection and contacts with customers. For these reasons it was thought sharing would have to be in the form of aggregated statistics rather than unit records.

It was quickly established that private sector data would not have the depth of coverage of sources such as the census and also established that this was no barrier to its value to Beyond 2011 in the eyes of ONS.

It was felt that in many locations the population/customer bases changed little over periods of many years and data had a long shelf-life. Conversely, in some locations change could be dramatic. There seemed merit in focussing both ONS research and the use of private sector data on such ‘difficult’ areas.

There was a desire that data be released with much less delay. Some public data sets can be released up to a couple of years following there data of collection. A system that published (less accurate) provisional data much more rapidly, with a process for publishing revisions was of much interest, although too many or too frequent revisions were to be avoided.

Users had a requirement for a relatively limited set of data. Geographic precision was of great importance but the range of topics could be much less than published in the traditional census. Population, age, households, families, housing counts were of more interest while some of the detailed cross-tabulations of little interest.

There was discussion of the various public data sources that might contribute to Beyond 2011. While one option is for ONS to combine all these into a set of statistics similar to a census people felt they might prefer to have more access to the individual data sets.

Sharing knowledge was thought to be a fruitful option. People felt that analysts in industry had developed skills in processing sparse data, merging data sets and the like and so there was much that could be offered to ONS in this area. There was some interest in a scheme of secondments of ONS personnel in industry and vice versa. Industry felt they would learn much more about public data and ONs analysts would develop an understanding of how decisions industry can at time of necessity be fast and based on data best described as ‘what we have available’ rather than precisely documented and refined data models.

John Rae, 14 October 2011