

Tracking a Decade of Changing Britain

How businesses are benefiting from early use of the Census and other Open Data

Seminar Report by Peter Furness and Peter Sleight

Introducing The Open Data Opportunity

This was the third seminar organised by the CGG and having the 2011 Census as a core theme. This time we have broadened the scope to focus on how the Census is being applied, whilst, at the same time, including other sources of Open Data

Open Data, of which the Census is a part, is a big deal. A recent report from McKinsey and Company shows how open data can help create \$3 trillion a year of economic value across seven sectors globally. The seven sectors are Education, Transportation, Consumer products, Electricity, Oil and Gas, Health Care and Consumer Finance. The potential value is shown to be divided roughly between the USA (\$1.1tn), Europe (\$900bn) and the rest of the world (\$1.7tn).

In this era of Big and Open Data, companies and individuals have access to vast amounts of data delivered through the web and through corporate and government databases and systems. Companies are using this data in powerful analytical tools to operate and market more efficiently and to identify and exploit new market opportunities. As individuals we use this data in almost every aspect of our lives, from education to health, to online shopping, digital entertainment, vehicle navigation, house buying and financial planning - a vast and growing list.

Census 2011: New Opportunities, Analyses and Future Options

In the first session of the day Professor David Martin (University of Southampton) gave a wide ranging review of the 2011 Census and the new opportunities it is opening up, as well as a summary of the 'Beyond 2011' options.

A key message for census users is that 2011 is likely to be seen as the best recent census. A huge effort has gone into producing a nationally consistent design and an evaluation of both the process and the results. This includes, for example, an automated zone design based on the 2001 methods but with essential maintenance, which provides a high level of geographical consistency for making 2001-2011 comparisons. As examples of the types of 2011-2011 comparisons, Professor Martin cited the age changes being seen in London and the South East. Additionally, the data is being released as Open Data through a variety of new delivery channels, thereby making the data easier to access by a wider user base.

An important development with 2011 is the creation of new workplace geographies which will overcome a limitation of 2001 where workplace data was output for the residential geography. In the new design, any workplace zone should have at least three postcodes and one hundred workers more than are in its largest postcode, whilst respecting constraints involving industry homogeneity and shape. The new workplace outputs should be available in spring 2014.

Professor Martin concluded by surveying the latest developments in the 'Beyond 2011' programme. Two main options have been published in the ONS consultation document. The first is an essentially conventional census but with collection mainly online. Challenges for this option will be coverage, non-response bias, follow-up costs etc. The second option will use administrative data to create

annual demographic counts supplemented with a rolling programme of 4% sample surveys for characteristics. The administrative data will be mainly health/DWP/HMRC/education sources but the complexity of the sample survey designs is likely to be a big challenge. Everyone is encouraged to respond to the consultation as soon as possible!

Realising the Benefits from the Census

Oliver Doerle of the Office for National Statistics (ONS) then presented on the work of the 2011 Census Benefits Realisation Team as well as bringing us up to date with the 2011 Census release schedule.

The Benefits Realisation Team has adopted a two-fold approach. Firstly, to make sure that the benefits of the census are realised, by raising awareness and encouraging people to use the outputs. Secondly to gather examples of the use of census data and to quantify the benefits obtained. This was achieved through a combination of surveys and direct engagement.

The surveys were of the commercial sector and local authorities with the aim of understanding how census data is used. The results will be used for the 2011 census evaluation and will also feed into the Beyond 2011 programme. In the case of the commercial sector a limited but positive response was obtained. There were thirty seven responses in total, with the majority from the market research community and consultancies, but also with some responses from companies in financial services, the media, consumer goods and IT. A wide range of census applications was found including market and consumer segmentation, forecasting and planning, survey sampling, investment appraisal and funding bids. In the market research sector alone, a best estimate of the annual benefit of the census to the market research sector is around £70million annually.

In the direct engagement, face to face meetings were held with a range of commercial sector organisations from retailing, utilities, financial services, consumer goods and other sectors. These meetings identified a wide range of census benefits in areas such strategic research, product development and positioning, store catchment analysis, audience segmentation, and many others, including the management of staff diversity. In addition, meetings were held with local authorities, the emergency services and the voluntary sector, identifying a further wide range of census benefits. For more information see the [Census Benefits web pages](#)

In conclusion Oliver gave a summary of forthcoming 2011 Census releases which include the remaining multivariate outputs (labour market and qualifications 11/2013; car ownership & travel 01/2014, alternative population bases (workday 31/10/2013 for OAs and 01/2014 for workplace zones), and UK outputs (from 11/2013). Microdata and origin/destination statistics will be released in the spring of 2014.

Oliver also encouraged delegates to contact the Census Benefits Realisation Team (benefits.realisation@ons.gsi.gov.uk) with examples of how census data has made a difference in their work. With agreement from the organisations, ONS would then publish some of these case studies on the census benefits web pages to educate others about the benefits of census data. He also encouraged delegates to respond to the Beyond 2011 consultation which closes in December.

Approximating Social Grade on the 2011 Census

Helen Lambert of GfK NOP then gave a detailed account of how Social Grade is being approximated on the 2011 Census. Social Grade is an important discriminator in commercial and social research because it differentiates on a range of attributes for consumption as well as attitudes. It is vital as a 'standard currency' for sample control and weighting and, for this reason, it is important to be able to profile geographic areas and niche audiences to make the census data relevant to Market Research.

Social Grade is a demographic segmentation whose definition is maintained by the Market Research Society and was originally developed for the National Readership Survey. Its computation on the 2011 Census has been a team effort between the CGG and ONS and involved building a Social Grade allocation model using the NRS sample for Q4 2010 and Q1 2011. This sample had both Social Grade and the relevant census-type variables which would be used to build the model, including Employment Status, Working Status, Tenure, Qualifications, SOC 2010. A decision tree methodology was used to model Social Grade from these variables and the resultant algorithm could then be applied by ONS to the 2011 Census.

The correspondence between modelled and actual social grade on the NRS sample is very good. Furthermore, on the 2011 Census it is possible to compare the modelled (or 'approximated') social grade distribution to that on the NRS as a further check on the quality of the modelling. The distributions are very close but with a slightly higher allocation to Social Grade D instead of E and a slightly lower allocation to grades AB in favour of C1 which is most noticeable in the 65+ category and most likely due to the allocation based on previous occupation. Helen showed some examples of maps social grade distribution at Lower Super Output Area level in London and these fitted very well with local knowledge of the areas concerned.

Approximated social grade on the 2011 Census has already been published in England & Wales to the Output Area level. The data will eventually be published at the equivalent level for Scotland¹. The approximated social grade data will have many valuable applications; for example, in geodemographic analysis, targeting local markets, sample design & execution for surveys, and as a link between the census and customer database analysis and survey research.

Open Data

Heather Savory, Chair of the Open Data User Group (ODUG) gave a wide ranging review of open data and its benefits and also on the work of ODUG with details of the results obtained by ODUG to date as well as challenges and next steps.

The Open Knowledge Foundation (OKNF) gives a definition of open data thus "A piece of data or content is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike." It is important, however, to point out that open data is only appropriate for non-personal information and where there are no security or legal concerns about release.

The benefits of open data include: transparency (holding government to account), better public service delivery, improved public sector efficiency, economic growth and efficiency, innovation and enterprise. The benefits of open public sector information to the UK economy have been estimated to exceed £1.8bn annually

¹ Approximated social grade data for Scotland was eventually published on 18/12/2013 (as part of Census Release 2C down to OA level). Also, the corresponding data for Northern Ireland was published in April 2013, so data for all countries of the UK is now available.

of direct benefits with wider social benefits in excess of £6bn. Internationally, the value of GeoServices alone (services which use open GPS data) are estimated to be \$150-\$270bn annually and growing 13% per year.

ODUG advises the Public Sector Transparency Board on public sector data that should be prioritised for release as open data, to the benefit of the UK. It is an 'intelligent customer' on behalf of the public and private sector for public data free at the point of use and for re-use. Also it champions the benefits of using open data and advises on National Information Infrastructure. The first members of ODUG were appointed in June 2012 and are volunteers representing different sectors of the user community (large/small businesses, local government and the wider data community). ODUG has adopted a fully open and transparent approach (e.g. with minutes published) and maintains a Data Request mechanism (at www.data.gov.uk) and Data Request Roadmap to track the status and progress of requests.

Data.gov.uk now has over 10,000 datasets available as open data. ODUG has already achieved a number of notable successes; these include for example HM Land Registry Historic Price Paid data, and Ordnance Survey Public Rights of Way data. In addition, ODUG has managed to get Government to look at options for an open National Address Dataset, to commit to release the Register of UK Charities, and to consider releasing Historic UK Met Office Data, the VAT Register, DVLA Bulk Data and the Police National Stolen Vehicle Data

There are many challenges. Some are cultural, although public sector data should be open by default the associated legislative landscape is complex. Clarity is required on requirements for the commissioning of new data, the Government's policy on data privacy, and challenges to progress caused by the skills gap. There is also the difficulty of making the case for open data as a driver for growth (metrics can be difficult and evidence will only emerge slowly as data is released. Competing policy agendas in a time of austerity do not help).

Heather concluded by looking at the next steps for ODUG. The Year Two focus will be on five main areas. Firstly, the National Data Infrastructure and Core Reference Data. Secondly, the improving of the 'data requester experience'. Thirdly, the building community engagement. Fourthly, the thematic approaches to data release; and, finally, the production of more open data case studies.

Introduction to Geodemographic Classifications

Peter Sleight kicked off the afternoon sessions with an introduction to "geodems", giving a brief history of the geodemographics industry over the past thirty five years (in which Peter, himself, has been a key player). He emphasised that knowledge of previous classifications is important for understanding future ones.

Geodemographics and the New Economy

Simon Whalley from Beacon Dodsworth gave the first of the supplier-focused presentations. He described the new P2 People & Places classification, with a particular emphasis on its application to understanding how the recent recession has impacted the population.

The classification was developed by clustering the UK's primary urban areas economically on the factors that affect the performance of a city – for example, the rate of business churn, the level of knowledge-based intensive services and the incidence of most highly skilled workers. Simon used maps of various parts of the UK to illustrate the results; including Cambridge – which is part of the "Qualified Growth" cluster, because of its high levels of qualified people, many high-tech companies and potential for growth.

The next stage of development will be to build P2 People & Places inside the economic clusters. Beacon Dodsworth has been working with Professor Peter Batey and Dr. Peter Brown (formerly authors of Superprofiles, a classification from 1991; also involved in the P2 People & Places classification in 2001) from the University of Liverpool.

Beyond 'lazy' Geodemographics – Streetwalkers, The Internet, and Freedom of Information

Next up, John Rae described CACI's very different approach to geodemographics which was used for building the new Acorn system. In this radical new methodology, the segmentation structure is created first, and then small areas are assigned to segments using various methods and data, chosen according to the circumstances.

The data included small area statistics created from Open Data, administrative data, and various private sector datasets. In particular, as John emphasised, the new Acorn was built and tested prior to the release of 2011 Census data. Subsequently, the effect of adding census variables was tested and it was found that these variables made little difference to the power of the new classification – so the Census is "nice to have, but not essential".

The Next Generation of CAMEO UK: What to Expect

Paul Kennedy, from Callcredit, then outlined the approach that his company is taking to rebuilding the CAMEO segmentation. Callcredit has taken a very different approach to that of CACI, by using 2011 Census data as the backbone for their new product. Paul explained the importance of the 2011 Census for understanding migration and ethnicity, as well as for tracking population change.

He went on to explain that, in addition to Census data, there will be a diverse range of behavioural, lifestyle and survey datasets used in building the new CAMEO – for example, in creating colourful segment descriptions and for ongoing "dynamic updates".

How the 2011 Census Data Supports SoLoMo (Social Location and Mobile) Solutions

Andy Bell from Pitney Bowes Software was the last of the commercial presenters and he focused on explaining how 2011 Census data supports Social, Local and Mobile (SoLoMo) solutions. He explained how consumers now expect brands to deliver personalised experiences, using more precise locational and behavioural information. Census and geodemographic data are used to enhance these services by providing the context about people, according to where they live.

Pitney Bowes plan to merge these sources together with a variety of other data – such as consumer and business locations, types of areas and points of interest – for location-based profiling and targeting.

The 2011 Output Area Classification for the UK

Chris Gale from UCL gave the final presentation of the afternoon, this time from the academic sector. He presented on the 2011 Output Area Classification (OAC) which has been developed in partnership with ONS. The 2011 OAC has been developed using a broadly similar approach to that used for 2001 OAC and relies

only Census data - in fact it is the only geodemographic discriminator to be based solely on Census data.

Chris explained that a larger number of variables were available for 2011 than for 2001 and this has resulted in a classification that contains a greater number of categories at each level. He illustrated 2011 OAC very effectively, using a large number of colourful maps and charts. One such map was particularly memorable, comparing two different areas - most of the people who live near to Liverpool FC are Hard-Pressed Households, while those near to Chelsea FC are mainly Cosmopolitans!

Summary of Classification Developments

Peter Sleight concluded the afternoon sessions by giving progress reports from other developers who are currently building 2011 geodemographic classifications, including Personix from Acxiom, Mosaic from Experian, Censation from AFD and Sonar from TRAC. Full details of this and all the afternoon presentations can be found in [Peter's full report on the new generation of geodemographic classifications](#)

Further information about the seminar, including speakers' presentation slides, is available from the Market Research Society by [clicking here](#).

Also via the [Geodemographics Knowledge Base](#)