The New Generation of Geodemographic Classifications

by Peter Sleight, Target Market Consultancy

Based on the MRS CGG seminar:
'Tracking a decade of changing Britain'
Afternoon session: 6th November, 2013

The presentations summarised in this report are downloadable from:

https://www.mrs.org.uk/intelligence/cgg/events/trackingadecade

1. Introduction to Geodemographic Classifications – Peter Sleight

Peter introduced the topic of geodemographics, which he described as 'classifying people by where they live'. It involves an analysis of Census data (usually – but we will hear about alternative approaches and additional data this afternoon). Geodemographic classifications generate 'typologies' of neighbourhoods; postcodes are then linked to these neighbourhoods, so addresses may be classified.

It was the MRS Conference of 1979 that effectively introduced geodemographics to the commercial world. Three BMRB executives presented a paper to the conference in which they demonstrated that a ward-based classification, then known as CRN (which had been developed by Richard Webber, then at the Centre for Environmental Studies) produced a surprising level of discrimination when used as a sample frame for the Target Group Index. They found that CRN could distinguish, for example, between readers of the Guardian and the Telegraph more effectively than the 'traditional' indicators (social grade, age or education). This was because of the locational element of geodemographics; the Guardian was generally read by Inner-City high-status individuals, while Telegraph readers were predominantly Suburban and Rural high status. (This was only one example among many).

The Marketing and Media world got very excited about this. So CACI, which had been operating for a few years in the UK selling its 'SITE' census analysis system on a bureau basis, recruited Richard; his CRN system was rebranded ACORN, and the UK geodemographics market was off and running.

Peter then outlined the market proliferation that took place in the 1980s (companies such as PinPoint Analysis, with PiN and FiNPiN; CCN with MOSAIC; CDMS with Superprofiles, Infolink with DEFINE (Infolink was taken over by Equifax in 1994).

Peter then 'jumped a decade' to show the line-up of geodemographic suppliers post 2001 Census; see slide 1 below, for the list. Some of the names have since changed; Claritas was acquired by Acxiom (who dropped the PRIZM name, and substituted Personicx). EuroDirect changed its name to Callcredit, but kept the CAMEO brand name. CCN had changed its name to Experian in 1997. When we review the current (2013) line-up of geodemographic providers and their brands, we'll see a remarkable similarity from those of a decade ago.

Slide 1: Origins of Geodemographics

MRS Conference 1979; ACORN burst onto the scene
Market Proliferation in the 1980's
Geodemographic classifications post 2001 Census

• AFD/ Data Talk	Censation Open Data
• ONS / Academic	Output Area Classification (OAC)
• Beacon Dodsworth	P², People & Places
• CACI	ACORN
• Claritas	PRIZM
• The Clockworks/TRAC	Sonar
• EuroDirect	Cameo
• Experian	Mosaic
• GeoBusiness	Atomic
• ISL	RESIDATA Lifetypes
• Streetwise Analytics	Likewise

1 Tracking a decade of changing Britain – 06/11/2013

2. Geodemographics and the New Economy – Simon Whalley, Beacon Dodsworth

Simon started with two questions:

- Who feels they have been affected by the recession?
- Who feels that the economic situation is now improving ?

He illustrated this with two headlines from 2008/9 –

'Three weeks that changed the world', and

'Everything you want to know about the bank crisis', followed by a graph of UK Economic |Performance, illustrating steady growth to 2007, followed by collapse thereafter.

He illustrated that the economic recovery has wavered in the months from the start of recession.

He asked 'What about the future?'

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 P^2 People & Places classification in 2001) from the University of Liverpool. The team was focussing on:

'Tracking a Decade of Changing Britain'

- How well are areas doing at present?
- How well are they likely to do in the future?

The 'Centre for Cities' Research was:

- Looking at Primary Urban Areas within the UK
- Factors that affect them all (past, present and future)

There are 64 Primary Urban Areas in Britian

- Built up from Local Authorities
- Represent large towns and cities within the UK

This was illustrated on a UK map, which showed the 64 'Primary Urban Areas'.

Then Simon went through a process which reviewed 'What affects a City?', using variables such as 'density of an area' 'retired population (65+)', 'qualifications, level 4 and above', and 'economically inactive rate'; there were eleven such variables. Each set was illustrated by maps, showing the pattern of these variables in Primary Urban Areas. (See Simon's presentation for details.)

The culmination of this exercise was an analysis of the factors that either led to success, or 'struggle'! To be successful, areas needed:

- High level of qualifications
- Lots of innovation
- High tech companies
- Large potential workforce

Whereas, to struggle, areas with:

- High Public Sector employment
- Low qualifications/skills
- Manufacturing companies
- Small potential workforce

This information was put into cluster analysis to produce 6 clusters of Local Authorities; 'Urban Clusters'. As an illustration, Cluster 2 is named 'Regional Growth Centres'. Significant Positive Variables are:

- Business Churn
- Density
- Economically Inactive Rate
- Highest Level Qualification Rate
- 'Knowledge-Intensive Based Services'

Significant Negative Variables:

- Retired Rate

Areas within Cluster 2 would have a good chance of recovering quickly.

This was illustrated on a map; Cluster 5 was also illustrated.

Simon then completed the picture, by dealing with the remainder of the UK; by creating three Rural clusters, Beacon Dodsworth thus ended up with 9 distinct types of area; an economic classification of the UK.

Beacon Dodsworth's geodemographic classification, P^2 People & Places, was then introduced and described. As Simon explained, the new version of P^2 will be 'built inside the economic clusters'; he illustrated this with reference to Cambridge. (P^2 is a hierarchical system with three tiers; these are each ranked in order of affluence. There are 14 'Trees', 41 'Branches', and 157 'Leaves'. See Beacon Dodsworth literature).

In conclusion, Simon promised regular updates to areas of most change; stressed the economic dimension to this project; and the link to P^2, the geodemographic classification.

3. Beyond 'lazy' geodemographics- streetwalkers, the internet, and freedom of information – John Rae, CACI

John explained how the new Acorn has been built without using any Census data; given that 'Census-like' data has been used, this is nothing short of revolutionary.

The rationale for this approach was:

- ONS may not field a 'traditional' Census (i.e. with small-area attribute data) again;

  ONS' 'Beyond 2011' team has been investigating using admin. data (inter alia) to build future Censuses;

So CACI investigated building Acorn using small area statistics created from Open Data, administrative data, and private sector datasets. The team also took a fundamental review of 'how geodemographic classifications are built', over four decades of geodemographics. activity. The 'standard rules' have remained the same:

- take Census data and other national data
- apply a single statistical method to build a segmentation
- assign all households to types at the same time
- apply the same data to every postcode or household

CACI reviewed this, and decided to take a different approach, based on datasets that they had access to, or could acquire. First they defined the structure of the new Acorn, then they assigned entities, using many different methods according to the circumstances. They used different algorithms and different data processes. Different algorithms were used for:

- regional methods
- newly-built housing
- manual classification
- housing for specific categories of occupants

Also, input variables were derived from combining multiple sources; e.g. for local-level housing type and tenure:

- Land Registry

- Registers of Scotland
- Public registers of Housing in Multiple Occupation
- Freedom of Information (FoI) requests to Local Authority Districts
- Zoopla property portal
- CACI lifestyle database
- CACI high-rise dwellings database

For local-level family structure, occupation and affluence:

- CACI names and addresses
- Credit data
- 'Emma's diary' children's database
- DWP claimant data
- CACI lifestyle database
- UCL ethnicity imputation
- Company Directors
- Shareholders
- Students

For high-rise buildings:

- Identify likely location
- WALK THE STREETS!
- Create a database of addresses:
    . Social high-rise (10+ storey)
    . Social mid-riser (5-9 storey)

Having collated these datasets, CACI built Acorn without 2011 Census inputs. They then linked Acorn to research surveys and an insight test-bed, and optimised the solution across a large range of topics. They finalised support material for the solution before the 2011 Census was published. CACI considered adding Census data into the solution; but observed that this made no difference to the Acorn structure. So they launched the Acorn solution for England and Wales in March 2013. The solution has 5 high-level categories, 17 mid-level groups, and 59 detailed types.

John concluded by asserting that 'the usual questions no longer have the same relevance 'when building a geodemographic classification; that, if you 'make more effort', that is:

- FoI requests and, if necessary, appeal to the Information Commissioner
- Actively collect data, if necessary by manual technique
- Access Public Registers and Open Data;

then, you can optimise the solution. Finally, 'The Census is nice to have, not essential'.

4. The Next Generation of CAMEO UK: what to expect – Paul Kennedy, Callcredit

Paul's presentation was subtitled 'Tracking a decade of changing Britain' – how Callcredit are using Census and open data to develop our CAMEO consumer classification system.
Paul went on to highlight a number of dramatic changes over the decade:

- The proportion of the population 65 – 74 who are economically active almost doubled to 16%
- England/Wales showed a population growth of 7%, half due to immigration (highest ever level of immigration)
- Dramatic rise in number of 'other household' types, e.g. multi-generational families and unrelated people sharing
- The white ethnic group fell by 5%
- 1 in 3 London residents are non-UK born vs. 1 in 20 for the North-East

In an ethnographic survey of London, Manchester and Birmingham (1500+ families surveyed), it was found that the average home has 10 'connectable' devices, of which 6 are connected to the internet. Amazingly, 1 in 3 families use their technology to communicate within the home! (whatever happened to just shouting up the stairs!).

Paul positioned Callcredit's product development against a background of 'Big Data', open data, and huge technological growth; 'the next generation of segmentation is coming soon…..'

'Callcredit's CAMEO consumer classification system is being completely rebuilt for launch in Q1 2014';

- New modelling techniques layer intelligence from multiple levels
- 3 tiers of data block (individual, household, postcode)
- Built on a solid base of new UK Census plus actual data, surveys, transactional and lifestyle data
- Dynamically updated from 'Define' to reflect key life events and changes
- Unique vertical sector product literature
- Improved collateral with online interactive guides

Paul explained the rationale for using 2011 Census data for CAMEO;

- When rebuilding CAMEO, there were some important reasons why we used the Census as a backbone:
    - Migration and ethnicity: how would you understand this without the Census?
    - Tracking population change and therefore informing decisions about planning commercial operations.

Callcredit's 'Define' database was illustrated and explained; data inputs concerning:

Property and Insurance

Borrowing and Risk

Income and Savings

Technology

Vehicle and Insurance

Holidays

Interests

Shopping

Readership

'Empathy'

Neighbourhood

Demographics

These inputs provided both 'Individual-level colour', and 'dynamic updates' to CAMEO.

Paul then explained how CAMEO could be brought to life using diverse datasets; one could 'drill down' from Local Authority, to Super Output Area, to Postcode, Street, Household, and Individual.

Finally, Paul illustrated on a world map, the countries where versions of CAMEO are available, and additionally, where CAMEO consultancy was available.

He raised the issue of the uncertainty about the future of the Census 'beyond 2011', but that whatever happens, proxy datasets are becoming available that organisations will be able to use for planning and marketing purposes.

5. How the 2011 Census Data Supports Social Local and Mobile (SoLoMo) Solutions – Andy Bell, Pitney Bowes Software

In his overview, Andy would first explain SoLoMo; then show how Census data and the subsequent derived datasets (e.g. Geodemographics) would support SoLoMo. He would show how disparate datasets could be merged to support SoLoMo applications (both data to understand the consumer, and data to understand location); and finally, give an outline of Census data-based PBS SoLoMo solutions.

What is Social Local Mobile? It is the culmination of technological and behavioural patterns to have more precise location-based and behavioural-based targeting activities. Consumers now expect brands to deliver personalised experiences; and there has been a dramatic rise in volumes of data generated to inform these processes. Social media is driving innovation for everyone.

To put that into context, Andy showed a slide containing some impressive numbers:

- 45 billion Apps downloaded from App Store
- 1.4 billion smartphones in use in 2013
- $270 bn. p.a. potential value of Geo(?) growing at 30% p.a.
- 30+ billion pieces of content shared on Facebook per month
- 1.2 billion (mobile advertising grew 100% over 2012; 50% is Location Based Marketing)
- 100% increase in IPad ownerships 2011 - 2012
- 75% people that use location-based services

Plus more facts about Big Data.

The Result – Multi-level Smart Geo Fencing (to locate individual/devices: In Trade Area, In Neighbourhood, In Shopping Mall, In Store: focussing in on individual's locations, in order to secure relevant messages to them.

So how does Census data support SoLoMo?

- In Social media and mobile marketing, through Geofencing (as above);
- In Retail analysis
- In Traditional Outdoor advertising and marketing
- In Public Services

Merging disparate datasets:

(a) Merging Census and Geodemographics with other datasets:
e.g.
    - where people are travelling to and from
    - when they are going
    - what they are doing
    - implied consumer behaviour and lifestyle

(Derived from GPS and Wifi data)

(b) Points of Interest
- Business Location
- Neighbourhood Boundaries
- Parcel Boundaries
- 'Indoor Mapping'

Location-Based Profiling Results

In a test, PBS were able to identify 8.8% of mobile users as 'travellers', thus the travel agency can target them differently in related mobile applications or web sites.

Improving Customer Insight

Building up a detailed picture of a customer, from first browsing interaction, to multiple browsing, to first purchase, repeat purchasing, to a full customer profile. Then becoming 'networked advocate'.

6. The 2011 Output Area Classification for the UK – Chris Gale and Paul Longley, UCL

Chris explained that the current Output Area Classification, or 2001 OAC, is a free geodemographic classification of the UK; it was built using 2001 Census data and Output Area data. (Output Areas – and now Small Areas - are the smallest level of UK Census Geography, with roughly a quarter of a million areas covering the UK).

The main focus of Chris' PhD has been creating a new version of this classification (2011 OAC). Because of Census output timescales, so far the product had been produced for England/Wales only; the full UK product would hopefully be released in early 2014.

OAC 2011 is a 'general purpose' classification; but it has a flexible methodology so can be applied to specialised variants. It uses only 2011 UK Census data (use of other Open Data had been considered, but would have introduced too many uncertainties into the classification). The methodology will be fully documented when complete. It uses open source programs, and the code used will form part of a wider release of supporting documentation.

Chris gave a brief description of each of the eight Supergroups, in terms of the characteristics of residents. He then illustrated OAC 2011's hierarchy of Supergroups, Groups, and Subgroups; and illustrated that 2011 OAC is better that its predecessor at differentiating London. He also compared the OAC profiles of six major conurbations, to illustrate the differentiation between them.

186 variables were considered for inclusion, and after extensive testing, 62 were selected to create the 2011 OAC. Chris illustrated the correlation matrix, then showed bar graphs showing the variables for England/Wales that are above and below the national average for each Supergroup. He explained that this is the information that facilitates the naming of the Supergroups, Groups and Subgroups. He went on to illustrate 2011 OAC Pen Portraits, using the 'Student Communities' Group as an example.

Chris showed various examples of data visualisation, in mapping form. He referred the audient to the ESRC funded Retail Research Data website, which offers interactive maps of 2011 OAC for England and Wales. Other data can be downloaded, including Pen Portraits, Radial Plots, Bar Graphs, and other documentation and data.

Finally, Chris summarised 2011 OAC; the England/Wales classification will have a limited shelf-life, as it will be superseded once the availability of data for Scotland and Northern Ireland can be added, and a UK-wide OAC classification produced. He made the point that the open methodology and use of open source programs means that anyone with the right skills can either replicate the classification, or modify it for their own purposes (e.g. to create bespoke geodemographic

classification). Work is being undertaken at the University of Liverpool on updating 2011 OAC in due course.

7. Summary of Classification Developments – Peter Sleight

Due to delays in the provision of Census data (particularly in Scotland) we're not as advanced as we might have been in terms of product launches. However, the table below shows the classifications that have been announced to date (a few of which have just been described).

1. Acxiom	Personicx
2. AFD/Datatalk	Censation
3. Beacon Dodsworth	P^2, People and Places
4. CACI	Acorn
5. Callcredit	CAMEO
6. Experian	Mosaic
7. TRAC Consultancy	SONAR
Open Data	
8. ONS/UCL	OAC 2011

I have some information about the products which we haven't heard about today:

Personicx Acxiom spokesman: 'Our new version will be quite different from previous versions – specifically, we have built one solution that works across all levels of geography – postcode, household, individual – and now incorporates a code that denotes 'digital sophistication'. Personicx is being increasingly used within online environments for ad targeting and is core to our recently launched cross-platform global Audience Operating System'.

'Specifically, we didn't use any recent Census data within this new UK version – we carried out many tests but it did not improve the overall model performance. We think this will be a key differentiator within the marketplace – our prime input into Personicx has always been our own on/offline survey data and we now have actual (responder) data for over 20 million UK consumers at their current address. This means that the segmentation performs particularly well at an individual/household level, specifically with behavioural variables.

Censation 'Free with AFD name and address management software, therefore the most widely-distributed UK geodemographic classification'.

A neighbourhood classification at unit postcode level.

Uses over 600 variables from the 2011 Census, plus

- 80,000 face-to-face interviews p.a.
- Referenced to residential and transactional data
- Uses Land Registry data

Methodology treats OA-level Census data as an 'undulating terrain', rather than a series of discrete 'islands'.

'Will indicate relative affluence, lifestage and lifestyle of those who live there'.

Built by Tim Drye, Datatalk Solutions.

Mosaic 'A Brand New Segmentation to Reflect the Changed Consumer & Societal Landscape'. Our new Mosaic will be a brand new segmentation built with new data and methodologies, supported by new user visualisation tools and other

exciting new development to embed segmentation more easily into our clients' organisations.

'Data Depth & Quality'. There will be a focus on the depth & quality of the input data with new sources of actual data to produce a more robust and stable segmentation. Mosaic will optimise the balance of geodemographic and individual/household data so it is much more than a neighbourhood classification. The same Mosaic types are assigned to both individual households and postcodes and they are fully otimised for use at both levels.

'Interpretative Data – Online & Offline'. It will use a richer depth of interpretative data to paint a more detailed picture of who target audiences and customers are than ever before – both online and offline.

'Cross Channel Enabled'. The new Mosaic will be truly cross channel enabled through our extensive linkage data to ensure marketers can serve consistent messages to their target audiences across the marketing mix and Mosaic will become a common currency for target audience definition across traditional and digital channels for our clients.

'Exciting new Visualisation & User Tools'. Our new Mosaic will be accessible to organisations when and where they want it including on demand, on premise, online and managed services to deliver insight at the point the client needs it. Experian are investing significantly in interpretation and user tools to ensure Mosaic is easy to understand and use effectively.

Sonar A new version of Sonar will be produced once the Census data for Scotland is released. Other data sources in addition to Census include:

- Council Tax
- Land Registry Prices
- Police Crime Data
- DWP claimant data

and the new classification will look somewhat like the old Sonar in that clusters will be post-stratified into groupings by lifestage and affluence.

Sonar is owned and built by David Griffiths of TRAC Consultancy.

I have also heard about another classification that is in development that its author does not want publicised at present; so that's eight commercial products this Census decade, against ten last decade, plus OAC as an Open Data classification, both times.