



The Future of Classification Systems

This article will consider some issues around the sourcing of data and technological development and what impact these will have on the geo-demographics market in the future.

Much will change over the next ten years: globalisation, politics, technology, regulation, business, consumer attitudes to name but a few. To be a certain age in ten years time will be different from being that age today - or to what was being that age ten years ago. Ways of marketing and channels used will change. But what will drive this change? Client demand, Data, Technology? The answer is all of these – and then there are those we don't know we don't know yet – the Black Swans (Taleb).

Government data

Census data has long been a mainstay of the geo-demographic industry. It offers the most complete coverage of any national database and forms the basis of most of the current generic systems. However, it is only available at a (relatively) high geographic level (Output Area) – circa 125 households in England / 50 households in Scotland. Therefore systems that use Census data as their main building block can only estimate the characteristics of a given household from that household's Output Area data.

Further, it is not a dynamic source. In the UK it is undertaken once every 10 years (the last time being in 2001). Thus the data is now 7 years old. A lot can change in this time – especially at a micro level.

For demographic information, the Census is a treasure trove as its emphasis is on satisfying Public Sector requirements. However, it has never been used as a tool to gather more commercially oriented Lifestyle data.

Finally, will it be around in its current format for much longer? Much consideration is being given to the format of the Census. Will the next one (due in 2011) be the last of the current format? Will a more flexible, rolling model be adopted as in France and the US? Or even a population register based one as in Scandinavia? Whichever model is chosen, there must be concern that the data gathered will be of use to the commercial sector.

There are other Government datasets that contain household and individ-

ual level data (e.g. HM Revenue & Customs, Immigration, Dept Work Pensions). Unfortunately, most of these are currently unable to be accessed on either privacy or technical grounds.

That's not to say that in future there won't be any access to them (just look at what has been possible through the use of health data via Dr Foster). However, this is unlikely within the foreseeable future. (And even less likely given recent data loss incidents!)

Commercial data

The gathering and sourcing of personal data by commercial companies faces some major challenges.

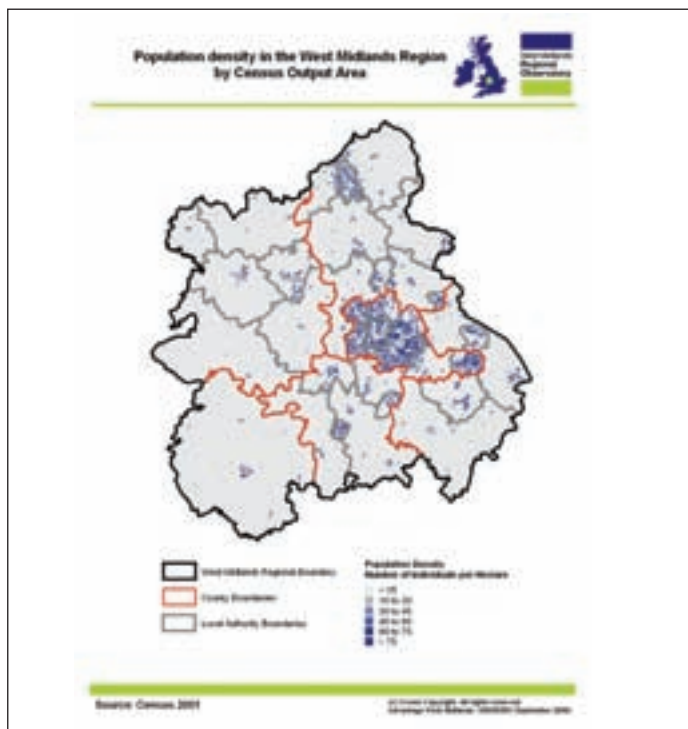
Response levels to traditional, multiple-section, paper-based surveys are dropping - not massively year on year but still dropping nonetheless. Such surveys will survive and be of importance – but will only be used for niche audiences (as will other channels such as phone and face-to-face interview).

The public are becoming aware and savvy as to the value and uses of their personal information. They are increasingly willing to part with their data if they see a direct return from it in terms of value. It's simply human nature.

Increasingly restrictive legislation is likely. This will favour the collection of data through established, known brands. ("Give my details to which company? I've never heard of you! How can I trust you? Goodbye!") Smaller and less established data gathering operations will suffer unless they have the backing of a recognised name.

Online data gathering is a rapidly growing area. In time it will become the dominant data acquisition channel. This will satisfy both the data giver as it will be their chosen channel and the data gatherer as it will enable cost effective collection.

Overall, the idea of "Data irrigation" will prevail - the consumer will provide small, relevant chunks of information as, when and how they choose on the



COA-population-density

basis that it will improve their choice/user experience. This is part of what is more widely referred to as the “Choice Generation”.

Classification Systems

In terms of geo-demographic systems, the idea of what constitutes a “segment” will evolve. We have seen the move from the traditional geographic area approach to (increasingly) households and individuals. This move will continue but evolve to include other dimensions such as actions (e.g. purchases, subscriptions), lifestyle events (e.g. births, marriages, deaths) and location

(e.g. commuting, at home, at leisure).

- There will also be new types of demands made of such systems.
- Bespoke segmentation systems (e.g. Volvo driving credit card users of Shell forecourts)
- Situational segmentation systems (e.g. commuting vs. at leisure)
- Localized / contextual systems (e.g. the Polish community in Reading)

The “geo” in geo-demographics will come into its own. It will no longer be just who, what, when and how - the “Where” element of such systems will become increasingly important.

The installation and use of GPS is already becoming commonplace on mobile phones. And the use of SatNav tools continues unabated. Some recent business deals (e.g. Nokia’s purchase of NavTeq and the proposed takeover of TeleAtlas by either TomTom or Garmin) give strong indications that these areas are converging.

So what does this mean? Well, “Time-Space Paths” will now become of increasing importance to marketers. Data capture will be possible as a change of location / circumstance / behaviour happens. This will enable dynamic, event-based segmentation systems.

Consider this. What you are thinking about and how you feel at different parts of the day and in different situations affects how pre-disposed you will be to receive and/or consider a particular marketing messages. You will be more appreciative of a work related message when driving to work than you would a personal one. Similarly, a work related message will not catch you in the right frame of mind when playing with your young family or watching a game of football.

The development of dynamic based segmentation systems would enable the correct message to be presented to you via the correct channel at the right time for you to receive it.

So that’s data sources. Let’s now consider how the actual data will be gathered and processed.

“Technology”

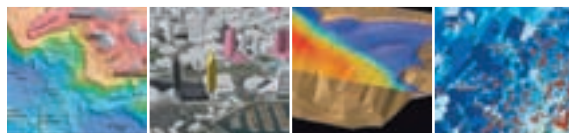
There will be ever more innovative ways to collect data; and ever greater volumes of it. For example:

UNIGIS

Study for a postgraduate qualification in GIS by distance learning

With over 16 years of experience presenting distance learning courses to professional standards, UNIGIS offers you access to the premier international network of Universities in GIS education.

Our courses meet the learning needs of busy professionals, or those seeking to enter the GIS industry. We support you with personal tutors, on-line help and optional residential workshops. Our courses are assessed by coursework - there are no examinations.



WE ALSO OFFER:

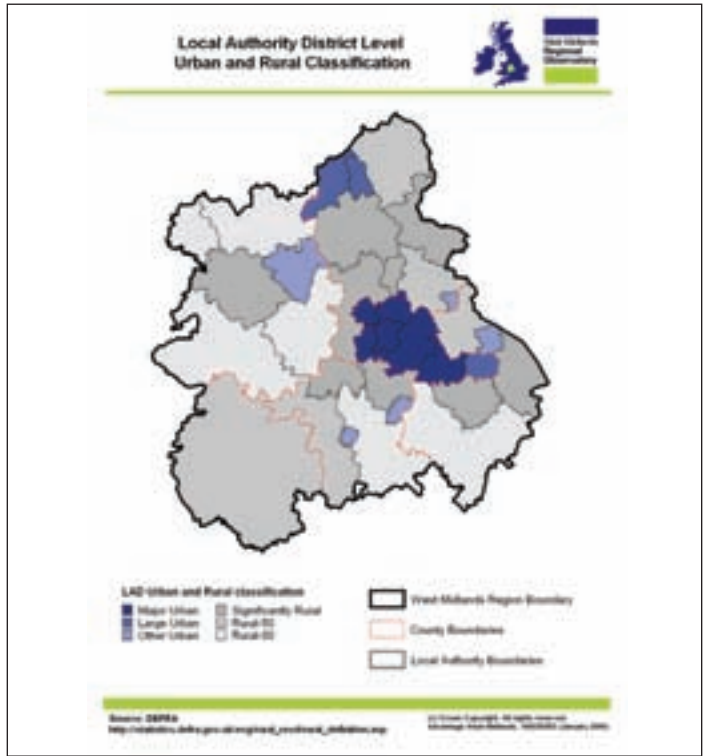
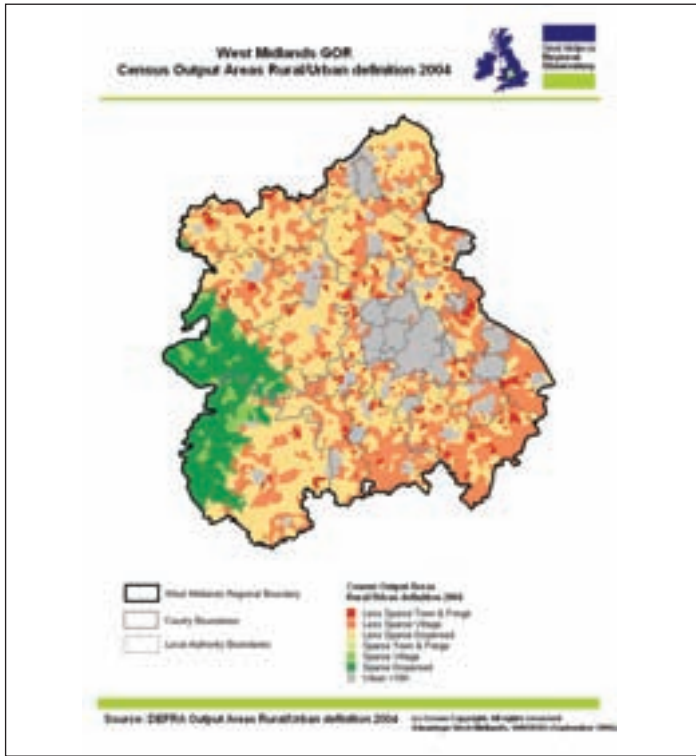
- Flexible entry requirements
- Specialist pathways in GIS, GI Science, GIS & Management and GIS & Environment
- Course modules supported by key textbooks, software and on-line resources
- Flexible study options - full distance learning or distance learning plus residential workshops
- No examinations - full continuous assessment, plus credit for prior learning or experience
- Networking with an international community of GIS professionals

Find out why the UNIGIS postgraduate courses are so successful: call +44 161 247 1581, fax. +44 161 247 6344, email unigis@mmu.ac.uk, or visit our web site at <http://www.unigis.org/uk>



Educating GIS Professionals Worldwide

www.unigis.org/uk



COA-rural-urban

LAD-Urban-Rural-Classification

Transactional data: We are all aware of the use of such data by retailers like Tesco. However, the use of RFID (Radio Frequency IDentification) in smart cards (e.g. Transport for London's Oyster card) offers a new twist on this idea. They will capture transactions (journeys, purchase, cash withdrawals) of users as they roam about. This will create a picture that, when aggregated, will show distinctive patterns of use that can then be related back to demographics and so enhance site location and service provision.

Unstructured data (Word, Powerpoint, pdf, email, websites): This will be increasingly tapped in to for information about individuals. A US company (MetaCarta) have developed a system that mines for geographic references – why not for other things (e.g. people)? If links can be established with an understood degree of confidence between these entries and other data then a vast new source of source of information can be utilised.

Visual data: Video, scanning, photos, drawings, paintings etc offer a new source of data. For example, video, long used for manual surveillance, is now being decoded by computers to identify the number of individuals in cars in dedicated multi-occupancy car lanes. How long before it becomes possible to identify individuals in a certain place at a certain time and then react to their presence in real-time (e.g. terrorist suspects)?

Social networks: Popular websites such as MySpace, Facebook and LinkedIn offer potentially huge sources of individual data. The exact extent of this potential is currently a great unknown! However, if you consider that the software lets people join and form online groups, list their interests and expertise, post text and pictures and all manner of data, communicate privately or openly with each other, and have all their postings tracked and analysed then you begin to see the opportunity for related and contextual data gathering through this medium. An aspect of such endeavour that will be of key interest to classification builders is that previously unidentified relationships will be revealed.

Technology will enable the above data sources to be fully utilised. For example, first we had access to the internet via mobile phones. Now, as we saw earlier, GPS has joined them. The next step will be the integration of some form of RFID smart card into mobile handsets. Once this happens the collection of the data we have discussed will become a formality.

But where is all this data that is being captured going to be stored and processed?

Grid technology, with its bolt-on devices offering limitless storage and its deconstructed processing, could offer an answer. However, this is currently a costly option and one that requires more development to be undertaken by the vendors to make it truly viable.

A future alternative could lie in nanotechnology - that is, based on single atom storage building blocks. IBM claim that such a medium could enable nearly 30,000 feature films or 1,000 trillion bits of data to fit on to a device the size of an Apple iPod. Larger devices would make the storage issue disappear.

Such technical advances bring the development of systems described above and the prospect of (near) real-time updates of classifications one step nearer.

“Impact”

The web has “democratised” access to data. We are part of what is named the “Choice Generation”. We can select who we want to initiate dialogue with, when and through which channel more than ever before. The age of mass communication is over. Instead, the consumer has the upper hand in defining what interaction should take place.

The internet has undergone various stages of evolution. The current one, Web 2.0, is all about user content – lots of data but no meaning to it. As we saw above, sites such as Facebook are producing vast amounts of data but no commercially usable information. The next stage of web evolution, Web 3.0, offers the potential to address this.

The “Semantic Web” as it is known will make the meaning of online documents and data more accessible to computers. The goal is to apply computer-based reasoning to evaluate and filter massive amounts of complex, unstructured data. This will enable unseen relationships to be revealed that would otherwise have gone un-noticed.

The merging of geographical (location-based) information with the abstract information that currently dominates the Internet has been termed the “Geoweb”. It promises the creation of an environment where searches could be undertaken based on location instead of by keyword only – e.g. “What is here?”

Summary

We are at the dawn of an era where there will be a growing demand for bespoke and market specific segmentation. The geographic aspect of data will grow in influence and individual level classifications will complement neighbourhood ones. Technology will drive access to the new data needed as well as the means for processing it and distributing it.

However, all of this will depend on how privacy issues are addressed – but that is for a different article!

Stewart Dickson, Independent Data Consultant, stewart.dickson@hotmail.co.uk