Applied Geography 29 (2009) 289–298

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

Geodemographics, visualisation, and social networks in applied geography

Alexander D. Singleton, Paul A. Longley*

Department of Geography and Centre for Advanced Spatial Analysis, University College London, Pearson Building, Gower Street, London WC1E 6BT, UK

Keywords:
Geodemographics
Geoweb2.0
Neighourhood
Visualization

A B S T R A C T

This review begins by acknowledging the success of geodemographics as an important area of activity in applied geography. However, it then develops a critique of the conceptual and computational underpinnings of the approach, and argues that changes in data supply and online communication have rendered current practices obsolete. It presents elements of a new perspective, entailing: changes in the specification, estimation and testing of online geodemographic systems; adoption of consultative practices from online folksonomies; automated generation of pen portraits; and 'on the fly' visualisation of the outcome of geodemographic classifications.

Introduction

Geodemographics are small area indicators of the social, economic and demographic conditions prevailing in small areas, or 'neighbourhoods'. They have been deployed, with apparent success, for more than 30 years as applied geographic indicators in both tactical and strategic marketing indicators, by customer facing organisations in most advanced economies throughout the world. Prior to that, the lineage to geodemographic classification can be traced to the work of the Chicago urban ecologists in the late 1920s (see Brown & Batey, 1994). This review and interpretation anticipates the imminent demise of the general purpose geodemographic classification system. We reflect that current proprietary systems have not emerged through their inherent evolutionary superiority to all other depictions of neighbourhood differentiation (see Harris, Sleight, & Webber, 2005). Rather, we see today's geodemographics as, on the one hand, a lagged response to past impediments to effective data supply, computing and disclosure control and, on the other, disengagement by the academic community from the socio-spatial measurement task in ways that are generalised, timely, salient and safe to use (Kempf-Leonard, 2005). We reflect that current geodemographic systems cannot be considered to be explicitly spatial in design, estimation or testing, and that local context requires systematic consideration in geodemographic profiling (Fotheringham, 1997).

A final strand to our critique concerns stakeholder participation. We concur with the view that a more scientific understanding of the public requires improvements in the public understanding of science (Horlick-Jones, Rowe, & Walls, 2007), and anticipate attempts to foster a virtuous and self-reinforcing cycle of improvement in geodemographic classification through consultative exercises (Longley & Singleton, in press). Taken together, the enticing prospect is of a new

* Corresponding author. Tel.: +44 207 679 0581.
E-mail addresses: a.singleton@ucl.ac.uk (A.D. Singleton), plongley@geog.ucl.ac.uk (P.A. Longley).

0143-6228/$ – see front matter
doi:10.1016/j.apgeog.2008.10.006

geodemographics that is engineered for particular applications, that recognises the provenance of data inputs, that is scientifically reproducible, that has a wide stakeholder base and that utilises the best spatial analysis methods in a problem centred approach to market segmentation.

Background: geodemographics and representation

The world is of seemingly infinite complexity, and the representational view acknowledges the intrinsic human desire to bring order to it through classification (Longley, Goodchild, Maguire, & Rhind, 2005). The purpose of geodemographic classification might be summarised as the application of taxonomic principles to small neighbourhood areas (typically unit postcodes in the UK and Zip + four codes in the US) in the quest to develop scientifically valid intelligence about issues such as socioeconomic composition, consumption habits and attitudes to public service provision. The geographic dimension to such classification is inherently important from the end user perspective, since it is key to meeting the preferences of consumers of private goods and the needs of citizens for public sector facilities.

Socio-spatial classification of neighbourhood areas has a long history in human geography and urban sociology, which can be traced to the ecological dynamics posited by Burgess and the Chicago School in the late 1920s, and which developed through the social area analysis of Shevky and Bell in the 1950s to the factorial ecologies of the 1960s. This accumulated experience led to the emergence of commercial applications from the 1970s onwards, and the coining of the term 'geodemographics' by Richard Webber to describe the distinctive consumption habits to the 'average' residents of neighbourhood areas. This long and distinguished history has been reviewed elsewhere (see, for example, the excellent reviews of Bassett & Short, 1980; Brown & Batey, 1994; Harris et al., 2005) and will not be repeated here. However, looking prospectively, we question whether geodemographics is continuing to evolve. From the perspectives of both applied geography and commerce, we suggest that the commercial geodemographics industry, established in the late 1970s, is now reaching a point of stagnation in terms of system and that the tools and methods widely adopted in both commercial (Birkin, Clarke, & Clarke, 2002) and public sector (Longley, 2005a) applications are no longer cutting edge in terms of technology or application. We base this assertion upon a consideration of changes in the nature and configuration of digital data infrastructures that have created significant and growing opportunities for development of bespoke geodemographic discriminators, and argue that the infrastructure of e-social science further breaks down the distinctions between analysts and end users (Foster & Kesselman, 1999). Taken together, our view is that the research environment for creating geodemographic information systems has fundamentally changed, and that these changes have implications for greater empowerment of end users and knock on consequences for the business models that underpin the supply of geodemographic information.

If these changes can be taken on board, geodemographic classification will continue to support spatial problem solving by providing an organising framework for the generic characteristics of unique spaces (Burrows & Gane, 2006). Commercial applications will continue to focus upon the correspondence between geodemographic class and the consumption of goods and services, while in the public sector these representations of reality will increasingly be trusted by policy makers as tools for allocating resources effectively (Longley, 2005a). By extension, there will also be prospects for monitoring the dynamics of socio-spatial change in ways that move beyond current essentially 'black box' use of geodemographics in scenario development.

Historically, the main data source that has been used as input for creating geodemographic classifications has been decennial censuses of population, disseminated at enumeration district or output area level in the UK and census block level in the United States. In the majority of commercial classifications, these data are often supplemented with market research or commercial survey data, plus information from public sector sources such as county court judgements. The 'black box' nature of geodemographic classifications in a large part reflects past methods of data supply, as well as past controls upon access. In the UK, for example, issues of public sector cost recovery and control meant that prior to 2001 non-government and non-academic organisations were required to pay prohibitive licence fees to use Census of Population data: this led, at one point, to a duopoly in data reselling through geodemographic classifications. The roots to the term 'data warehousing' also refer to un-networked physical locations, with networking of software functions as well as data supply being uncommon until recent years.

By the turn of Millennium, access to most census data had become free of charge, in some settings (notably the UK) marking a refreshing departure from previous practices of cost recovery. Notwithstanding continued disquiet on a number of fronts (Arthur & Cross, 2006), many of the aspirations of 'free our data' campaigns have now been realised. Additionally, through Web 2.0 (O'Reilly, 2005), XML based data standards are increasing the availability of new and innovative geographic data sources. For example, the UK real estate website Nestoria (www.nestoria.com) now offers a free application programming interface which supplies property for sale data that are geographically referenced at street level. In terms of software, the open source movement which pushed forward the agendas of data standards and interoperability has also created a range of free-to-use public domain software packages capable of advanced statistical operations such as those required when creating geodemographics (e.g. R: www.r-project.org/). Together, these developments have led to the opening up of new approaches to geographical data management and visualisation using Geoweb 2.0 software architectures: see Haklay, Singleton, and Parker (in press). Viewed retrospectively, however, the combination of historic cost recovery regimes with respect to public sector data and the private sector origins of ancillary sources makes it no coincidence that the two largest commercial suppliers of geodemographic classification that have emerged in the UK also have interests in data warehousing and or credit information and service many of the needs of what Thrift (2005) describes as 'knowing capitalism'. These organisations have also become adept at assembling and enhancing public sector datasets (notably registers of electors) in ways that have not been pursued in academia or the public sector, thus providing enhanced competitive edge for private sector geodemographic products and services.

The precise composition of the database of variables that is assembled to create a geodemographic classification is heavily influenced by choice, convention and chance: the predilections and conscious choices of the classification builder, guided, for example, by perceived interactions between ethnicity, income and shopping choices; conventions (often vaguely defined) as to what constitutes an appropriate indicator; and chance factors governing what data sources happen to be available. In general terms, the creation of geodemographic summary measures proceeds using techniques of cluster analysis or profile matching, with the objective of identifying social similarities amongst variables within a dataset. In computational terms, the classification process is usually devised to maximise within cluster homogeneity while maintaining heterogeneity between clusters (Everitt, 1974; Gordon, 1999). This optimisation procedure is, however, guided by the transformations (such as normalisation) that are applied to the raw data scores, and the weightings that are assigned to particular variables. The outcome of this procedure is assignment of areas to clusters, on the basis of social similarities identified (or, perhaps more accurately, pre-specified) by the clustering procedure, in a way that is largely independent of locational proximity of areas that appear to share similar characteristics. Using this information, the classification builder then creates labels and descriptions which are designed to convey the salient characteristics of areas assigned to particular clusters. An example is shown in Table 1 for the UK ACORN ('A Classification of Residential Neighbourhoods': CACI Ltd., London) classification.

Conceptions of small area indicators

As the discussion in the preceding section implies, the field of geodemographics has often had an uneasy relationship with the conception and measurement of social constructs. Indeed it has be argued (Bassett & Short, 1980) that serious attempts to anchor neighbourhood classification to theory have not been made ever since the field observation that underpinned much of Burgess and Park's early conceptualisation of biotic and cultural strata in society was subsumed, first with the inductivism and post facto rationalisation of social area analysis (Shevky & Bell, 1955) and then with the mechanical reductionism of factorial ecology (Longley, 2005b; Saunders, 1995). It is clear that disjuncture between the conception, representation and measurement of the social constructs that underpin neighbourhood differentiation have endured until the present day. Moreover, concerns about the validity of geodemographics as social measurement have until recently polarised over time (see Longley, 2005b). To many, the theoretical frameworks underpinning the current geodemographic classifications remain rather weak. From the perspective of a marketeer, for example, Sleight (2004) acknowledges the weak central organising concept to geodemographics to be no more than that 'birds of a feather flock together'.

End users in the commercial sector apparently remain unconcerned with such considerations, as the benchmark from the perspectives of strategic or tactical marketing remain simply the 'equal share' response rate to untargeted messages, and repeat purchases of geodemographic systems provide evidence that they are apparently fit for improving upon this. More recently there has been some attempt at rapprochement between geodemographic practices with academic theory from the commercial (Webber, 2007) and academic (Burrows & Gane, 2006; Parker, Uprichard, & Burrows, 2007) perspectives, assimilating, for example, the notions of human and social capital formation (Putnam, 2000). In a similar vein, in drawing parallels between concepts of social stratification derived from sociological literature and labels used in the Experian Mosaic[1] typology, Webber (2007: 185) suggests that these 'incorporate language which corresponds closely to that used in the discourse on both globalisation and gentrification, and studentification', although this assertion is substantiated by only limited reference to the literature, citing only a single source of Atkinson and Bridge (2005). The shorthand labels and descriptions assigned to the different classes within any typology have without doubt been useful in promoting typologies, yet vivid description does not of itself substitute for transparency and unequivocal evidence of scientific rigour. Indeed, the user is provided with few clues as to the ways in which input data are weighted, severely impairing not only reproduction but also scientific questioning of the ways in which the clusters emerged from the underlying data. This is the more worrying when the data sources are themselves of unknown and possibly dubious provenance. Additionally, the detailed methods underpinning construction of the commercial classification are closed to detailed external evaluation.

Even in the light of post facto rationalisation of the conceptual foundations to commercial systems, it seems apparent that while geodemographic classification systems may be informed by intuition and the very extensive analytical experience of the person creating the classification, it is essentially the case that this experience is bolted onto prevailing sociological orthodoxies, rather than guided by them. There are worrying similarities between the criticisms of Shevky and Bell (1955) in relation to their post facto rationale of social area analysis. This stands in quite stark contrast to widely debated and well-honed conceptions of 'socioeconomic status' (Rose, Pevalin, & O'Reilly, 2001) that provide a much firmer and more explicit grounding for the classification of this widely used concept. However, the scope of the resulting concept is rather more restricted than geodemographic indicators in enlightening our understanding of neighbourhood lifestyles (Burrows & Gane,

[1] Mosaic is © Experian 2008. See www.business-strategies.co.uk/.

Table 1
A representative geodemographic system: CACI's ACORN Typology (source: CACI, 2005).

Category	Group	Type
Wealthy achievers	Wealthy executives	Wealthy mature professionals, large houses
		Wealthy working families with mortgages
		Villages with wealthy commuters
		Well-off managers, larger houses
	Affluent greys	Older affluent professionals
		Farming communities
		Old people, detached homes
		Mature couples, smaller detached homes
	Flourishing families	Older families, prosperous suburbs
		Well-off working families with mortgages
		Well-off managers, detached houses
		Large families and houses in rural areas
Urban prosperity	Prosperous professionals	Well-off professionals, larger houses and converted flats
		Older professionals in suburban houses and apartments
	Educated urbanites	Affluent urban professionals, flats
		Prosperous young professionals, flats
		Young educated workers, flats
		Multi-ethnic young, converted flats
		Suburban privately renting professionals
	Aspiring singles	Student flats and cosmopolitan sharers
		Singles and sharers, multi-ethnic areas
		Low income singles, small rented flats
		Student terraces
Comfortably off	Starting out	Young couples, flats and terraces
		White-collar singles/sharers, terraces
	Secure families	Younger white-collar couples with mortgages
		Middle income, home owning areas
		Working families with mortgages
		Mature families in suburban semis
		Established home owning workers
		Home owning Asian family areas
	Settled suburbia	Retired home owners
		Middle income, older couples
		Lower incomes, older people, semis
	Prudent pensioners	Elderly singles, purpose built flats
		Older people, flats
Moderate means	Asian communities	Crowded Asian terraces
		Low income Asian families
	Post-industrial families	Skilled older families, terraces
		Young working families
	Blue-collar roots	Skilled workers, semis and terraces
		Home owning families, terraces
		Older people, rented terraces
Hard-pressed	Struggling families	Low income larger families, semis
		Low income, older people, smaller semis
		Low income, routine jobs, terraces and flats
		Low income families, terraced estates
		Families and single parents, semis and terraces
		Large families and single parents, many children
	Burdened singles	Single elderly people, council flats
		Single parents and pensioners, council terraces
		Families and single parents, council flats
	High-rise hardship	Old people, many high-rise flats
		Singles and single parents, high-rise estates
	Inner city adversity	Multi-ethnic purpose built estates
		Multi-ethnic crowded flats

2006). It is difficult to conclude other than that geodemographic typologies still do rather little to provide, or even post rationalise, a conceptual framework to socioeconomic representation (Parker et al., 2007).

The computation of geodemographic indicators

In computational terms, geodemographic classification presents an important aspect of the tide of inductivism that has run through social science, which has gathered strength with greater availability of geographically referenced data and with developments in computer technology. Our own view is that this development has been overwhelmingly positive, and offers much enhanced prospects for formulating and validating theory through measurement and generalisation. This offers the prospect of a far closer rapprochement between inductive theory and measurement – but at the same time these very developments in computing, data collection and data access now undermine the business model which has led to a small number of classifications dominating the field.

Our view is that, in these changed circumstances, classification should be guided by application specific data which are intelligently sourced, validated and integrated based on existing theory applicable to specific application domains. Central to our prescription for future geodemographic classifiers is the fundamental principle which reconnects area classification to domain specific theories by revising their construction within carefully prescribed contexts; and as such challenging the assertion that a specific system should hold sway across all application areas. However, an important issue underpinning the development of any bespoke geodemographic systems is the extent to which appropriate data infrastructures can be developed for an application in any given sector. This is also true of some generic data sources: for example, UK Electoral Roll data are collected by the electoral registration offices in each separate local authority, yet no public sector centrally available source of the source exists, and as such, access to these data has to be negotiated separately with each custodian authority. The public sector also seems to perpetuate inherent disadvantage for itself *vis a vis* the private sector, by setting its face against the common commercial practice of utilising ancillary datasets to fill (or at least identify) the 'holes' in the register created by 'opt outs' from the publicly available edited version of the source. A further UK example is illustrated by educational data, which following devolution are now collected by separate agencies in each country. Additionally, within England, schools, further education, higher education applications and higher education admissions data are all collected by separate agencies. Because of these overall complexities in the UK data economy, many areas of the public services will not be informed by the best available data.

Before explaining how the construction of geodemographics might be re-orientated to build upon application specific foundations, it is important to consider how technological changes are shaping the specification, estimation and testing of a new generation of bespoke geodemographic discriminators. The catalysts to these changes can be broadly described in terms of developments in the computing environment and in the emergence of new local data infrastructures. The cumulative increases in the efficiency of data creation and storage, aligned with increased processing power and bandwidth connectivity have decentralised the capacity to create computer intensive models of urban areas away from large and expensive mainframe computers to anyone with a basic desktop computer client.

Despite this sea change in the creation and maintenance of geographic databases, geodemographic classifications are still the largely exclusive preserve of small numbers of skilled individuals who assemble datasets on a single computer for local processing – in the same tradition initiated by Shevky and Bell in the 1950s. This is a one to one relationship of data use in which the classification builder collects data, stores and processes them locally and uses a clustering technique to compile a classification, all on a standalone machine. End users may be unaware of: the range of sensitivity analyses or the number of different passes through the data that have taken place; the weights that have been assigned to particular variables; the validity or provenance of the data used to compile the classification; and the social constructs that the analyst had in mind when the classification was conceived (Longley & Goodchild, 2008). In commercial terms, this outcome is advantageous because it also preserves the intellectual property rights intrinsic to the development of commercial classifications. In application terms, it is very likely that many end users will not want to be overwhelmed by the detail of the classification procedure or the metadata associated with the constituent data. However, the spirit of recent open web-enabled computing is that the user should be able to interrogate the classification and ascertain its underlying data and assumptions as dictated by need. Instead, the prevailing orthodoxy fails to exploit any of the benefits arising from the creation of new local digital data infrastructures (particularly those pertaining to public service provision) brought by the Internet as a means of distributed information exchange, where data could potentially be updated on the fly from disparate sources; enabling live updates of classifications should the methods of construction be re-orientated to match this increased flexibility. There are many interesting data sources with high temporal resolution which are or have the potential to be made available in the public domain, and that could be integrated into an area classification service, thus refreshing census based information (Harris & Longley, 2002). For example, registers of GP practices are updated in real-time as people register and de-register from these local services, and with suitable disclosure controls this information could be made publically accessible at fine geographic resolutions (Gibin, Singleton, Milton, Mateos, & Longley, 2008). Quite proper ethical considerations surrounding the use of data on human subjects need not be used to frustrate the use of data. What is required is a new relationship between the provision and supply of geographic information and the processes of urban governance, that uses the geoweb to redefine effective participation, to better contextualise web-based models, applications and data, and to begin to create a cyber infrastructure that better enables two way interaction between citizens and government (see Sieber, 2008).

Additionally, the widespread availability of high capacity computing has enabled the development of a series of potential changes to the construction of geodemographic indicators. The issues that characterise the different priorities that might now be pursued are summarised in Table 2.

The majority of current geodemographic classifications fit a view of social science that patterning in the attributes of geographic spaces can be captured or measured at a global level. However, this is fundamentally at odds with a variety of developments in quantitative geography and computing, including geographically weighted regression (Fotheringham, Brunsdon, & Charlton, 2002), local indicators of spatial association (Anselin, 1995) and geographic data mining (Miller & Han, 2001). As discussed earlier, geodemographic classifications seek to simplify reality into understandable partitioning which

Table 2
Some different priorities that might guide geodemographic classification.

The more input data we include the more dimensions of reality we can detect.	The more input data we include the more noise we add, therefore, masking key discriminators.
Data normalising and scaling of input data should be maximised to reduce outliers (Vickers & Rees, 2007).	Outliers should be allowed in data scaling and normalisation as these tell us something interesting about reality, however, extreme values should be controlled for by variable weighting (Harris et al., 2005).
Clusters should be represented as discrete categories.	It is more logical to represent clusters with fuzzy boundaries (Feng & Flowerdew, 1998, See & Openshaw, 2001).
Classifications should be built initially at the largest aggregation, then broken down into finer groups (Vickers & Rees, 2007).	Classification should be built upwards by aggregating up from the level of individuals (Harris et al., 2005).
Input data should ideally come from sources with national coverage.	The inclusion of large or small sample surveys can add additional information into a classification.
Using principal component analysis is an effective means of creating classification input variables (Debenham, 2001).	Principal component analysis removes interesting dimensions of reality (Harris et al., 2005).
Input variables should not be weighted as this introduces personal bias into the construction process (Vickers & Rees, 2007).	Variable weighting provides an effective method of reducing outlier effects, thus creating more homogeneous groups (Harris et al., 2005).
It is preferable to use both *K*-means and the Ward methods of clustering as these have a long lineage of successful commercial and academic use (Everitt, 1974).	Modern clustering methods such as partitioning around mediods or genetic algorithms provide more efficient and effective means of cluster detection from within large multidimensional datasets (Brimicombe, 2007; Brunsdon, 2006; Feng & Flowerdew, 1998).
The clusters represented by a geodemographic classification are homogeneous across the UK.	Classification should be modelled to include regional differences (Debenham, Clarke, & Stillwell, 2003).
The variable choice, weighting scheme and clustering methodology used to create a classification are commercially sensitive and as such cannot be made available for public inspection.	For a classification to be scientifically reproducible and safe to use, the method of construction should be open to scrutiny, preferably appearing in the academic literature (Longley and Singleton, in press; Singleton and Longley, in press; Vickers & Rees, 2007).

through descriptions, pictures and labels give a good understanding of the constituent characteristics of areas. However, unlike geographically informed statistics, the classifications do not accommodate distinctive regional patterning of geodemographic classes. When current geodemographic classifiers are used to model a national population such as those neighbourhood types which nationally exhibit a high propensity to participate in higher education, then these types of statistics are reasonably logical – albeit, with the caveat that the clusters were created without being informed by local 'place effects' (Singleton, 2007). However, what is more problematic is when the global profiles of particular attributes are used to model the potential characteristics of local areas. For example, a model of higher education participation may target resources at a neighbourhood type which tends not to participate in higher education. This type of analysis will be geographically naive if the national picture is distorted by local circumstances as; for example, where a city academy school has ties with a local university, thus introducing a local deviation away from the national pattern of participation. Indeed, any successful locally targeted initiatives will affect outcomes and change the policy setting, and a task of geodemographics must be to accommodate these local variations.

In the past this type of simplification could be argued as necessary because organising and displaying multidimensional local statistics would be computationally intensive, and the tools for manipulating this information overly complex. However, with the advent of Google Maps and associated APIs for geovisualisation, these types of arguments break down. Using very simple technology it is possible to display a variety of data about local areas, through an interface which is familiar and simple to use: two examples of this type of implementation are available at www.maptube.org and www.londonprofiler.org.

Data dissemination and interpretation by stakeholders

A further area of contention in current geodemographic classifications concerns the openness of their construction methodology. When scientific methodology is closed, end users are not privy to the vagaries inherent in the specification, estimation and testing of the classification. Such situations were, however, the norm prior to the advent of what might be described as 'open geodemographics' – that is, the innovation of web-enabled classifications based upon turn of Millennium censuses. Of particular note in this regard is the UK National Statistics Output Area Classification (Vickers & Rees, 2007), which is supported by the Area Classification User Group of the Royal Statistical Society (see www.areaclassification.org.uk). This work is in the tradition of the open and free geodemographic classifications produced by the academic sector (Blake & Openshaw, 1994), with the addition of online help facilities. Such classifications fulfil the important criterion of scientific reproducibility. This property of openness can also be reinforced through appropriate public consultation exercises (Longley & Singleton, in press), in response to concerns in the literature about public accountability and the potentially deleterious effects of negative images upon place-based marketing initiatives (e.g. Goss, 1995). Our own attempts at 'classification through consultation' have entailed amending a niche classification of engagement with new information and communications technologies in the light of extensive feedback garnered through publicising the classification through a national (BBC) online news site (see http://news.bbc.co.uk/1/hi/technology/5256552.stm; see Longley and Singleton, in press). This approach is also potentially of use in ascertaining the need for updating of bespoke classifications of this nature.

Such developments move beyond the realm of post facto rationalisation of general purpose classifications. Instead, they have potentially far reaching implications for the widely recognised renaissance of interest in geodemographics from the public sector (Longley, 2005a; Longley & Goodchild, 2008). These changes are driven in part because of government pressure to demonstrate value for money (Ashby & Longley, 2005), and in part through the advent of new application areas such as community policing (Ashby, 2005; Williamson, Ashby, & Webber, 2006), health screening (Farr & Evans, 2005; Shelton, Birkin, & Dorling, 2006), analysis of educational attainment or participation (Batey, Brown, & Corver, 1999; Butler, Hamnett, Ramsden, & Webber, 2007; Singleton and Longley, in press) and regional planning (Batey & Brown, 2007). To date, the response of a number of established geodemographic providers has been, with some exceptions, to re-badge existing consumer classifications, and to seek correspondences between geodemographic classes and different attitudes and behaviours towards public services. Correspondences are, typically, sought by coding up large scale public sector services according to geodemographic category and applying these at the local level to uncover public attitudes to service provision and to devise appropriate interventions.

There are several objections to this approach, beginning with the issue of whether socio-cultural similarities in the consumption of private goods and services should map onto patterns of public attitudes towards service provision (Castells, 1973). Public services which are consumed collectively and public obligations towards service providers such as the police are also often discharged collectively. Why should a particular cocktail of variables and a weighting scheme that might be suitable for predicting aggregate shopping behaviour be central to understanding local attitudes towards community policing, preventive health care or school choice? To some extent, the reason why existing geodemographic solutions are still used is because the public sector fails to unlock the value of sector data collected at the local level. The consequence is often reliance upon crude univariate measures (such as local uptake of free school meals in education applications) or, at best, multivariate measures such as the Index of Multiple Deprivation (CLG, 2007). The market is clearly ready for more sophisticated area classification methods. As such, in the next section we will present a new model for area classification that represents a unified approach to input data specification; variable estimation and weighting through the development of online web interfaces; and finally; testing, both through consultation with the public, and through the development of other modes of stakeholder involvement that move beyond those total weighted deviation type assessments demonstrated by Webber (2004).

A new perspective

The geodemographic model presented in this section is a vision of how neighbourhood classification might be structured and driven by the best data, rendered accountable to different stakeholders, and used to create near real time representations of reality. Each of these characteristics is indicative of best practice in applied geography. Bowker and Star's (1999: 11) notion of classification as 'a spatial, temporal, or spatio-temporal segmentation of the world' is useful as it recognises inherent subjectivity, while also acknowledging the need for flexibility in accommodating temporal shifts in our understanding of the nature of organisation. Such inter-subjectivities are already acknowledged in many Internet representations: Weinberger (2007), for example describes the creation of user defined folksonomy classifications, which evolve to meet multiple informational requirements (see also Vanderwal, 2007). Folksonomies may also evolve over time through the use of 'tags' to describe photographs that have been posted on websites such as flickr (www.flickr.com). In a similar vein, geodemographic classifications could be recast as (partially or even wholly) user-generated representations of society. However, unlike the Internet folksonomies created by and for online communities, geodemographic classifications have hitherto always been created by small numbers of developers, who may be physically and socially remote from the stakeholders that use them and the people that are represented by them.

Moreover, an apparent legacy of past attitudes to cost recovery and restrictions in the availability and supply of data has been the promotion of 'one size fits all' classifications for the full range of purposes, with no right of appeal of individuals or communities against the categories to which they have been assigned. In contrast to this backdrop, our prospective vision for geodemographics embraces the practices of folksonomy, giving users access to an online service where they can create classifications which are tailored for particular purposes and which are safe to use.

There are a series of changes in technology that will enable the creation of such a service, not least developments in web browsers which enable technology such as DHTML AJAX or Flash. These open up the opportunity to create more complex user interfaces akin to existing desktop applications. Using such systems, it should be possible to build a geodemographic classification from scratch using a palette-based interface, from which a range of input variables could be selected. Users would also have the ability to set weights for individual variables, perhaps through the addition of slider bars. In ways akin to graphics software, where prior to applying a specific filter on an image it is possible to preview the settings selected, it should be possible to 'preview' an area classification. Perhaps an output could be created for a small subset of the data in a small preview window, thus providing the ability to readily gauge the sensitivity of outputs to inputs in the classification building process. A related group of technologies that are essential to the development of this service are the data sharing protocols based on XML which allow data in standard and structured formats to be incorporated into third party applications. For example, XML streams of census data could be combined with property information available from aggregation websites such as Nestoria (www.nestoria.co.uk) that provides KML files generated by search results of property for sale or rent. This will require the support of a more integrated approach to data sharing such as the adoption of standards, and further denuding the barriers that prevent the ready dissemination of geographically referenced data. A variety of metadata relating to source, integrity, scale, temporal currency and sample size would also be required. This will be absolutely essential if the service is to be used by non-experts, as this information could be used to provide a warning system about how data can be safely conflated and concatenated, or at a minimum warn as to when specialist advice is required.

A range of methodological variants have been incorporated into leading industry geodemographic classifications, sometimes with attendant claims as to their superiority over open techniques. Our view is that, as with data, all methods should be comparatively evaluated, and that the sensitivity of method to output created should also be rendered intelligible to end users and stakeholders. Geodemographics currently borrow strength from the social similarities between scattered areas, and through clustering in attribute space they do not directly exploit locational proximity or strength of association between

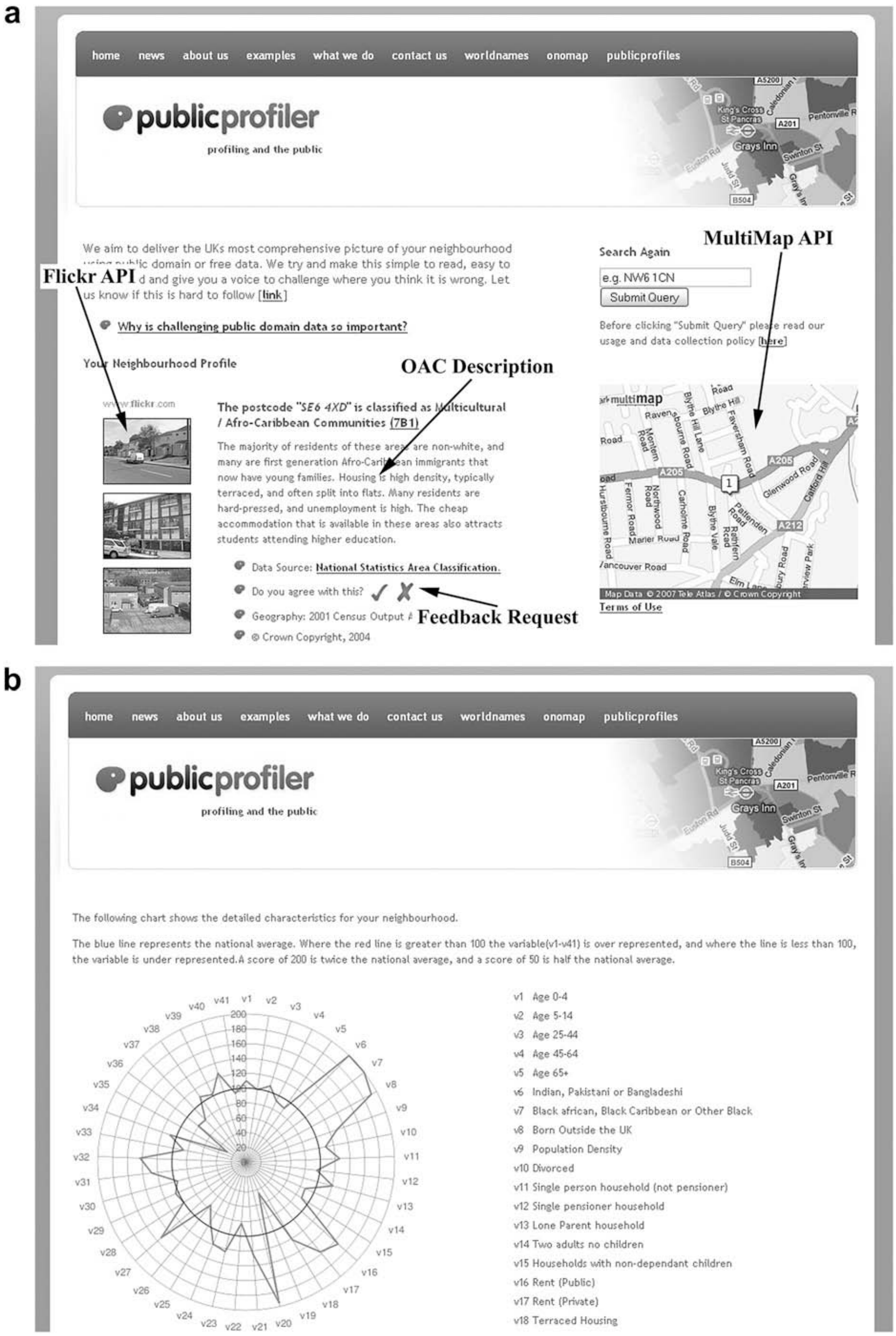

Fig. 1. The Public Profiler geodemographic website (www.publicprofiler.org) and geodemographic descriptions: (a) use of external APIs to visualise a typology; and (b) use of Google Charts API to graph statistics on a polar chart (OAC is the UK National Statistics Output Area Classification; flikr is a social networking site in which users can create shared pools of photographic images, and the Multimap site provides spatial referencing information).

areas. In some sense it is, therefore, erroneous to purport that geodemographics are spatial classifications, and the spatial autocorrelation of geodemographic classes across space does not emerge through specification in the system. As such it is proposed that further work is needed that might make classifications more sensitive to local conditions.

Additionally, if classifications are to be created on-the-fly, then the algorithms used must be efficient enough to enable the processing of a user specification within an acceptable time period. The majority of current geodemographic classification are created using *k*-means (Everitt, 1974); however, Singleton and Longley (in press) illustrate that this is unstable and requires repeated runs to convergence as their overall performance of a model depends upon the location of initial seeds. As such, further research is required to examine clustering algorithms which are less resource (time and processing) inefficient. An alternative solution for handling computationally intensive processing would be to create classification runs in parallel spanning many servers and utilising some of the advances that are emerging in the field of e-social science.

Once a classification is built and evaluated, the conventional approach is for the creator of the classification to describe it, using a range of methods which enable users to interpret the spatial arrangement and constituents of clusters. It is proposed that through the concatenation of a classification with a range of data, the propensity for specific characteristics could be generated as simple tables and charts, perhaps using a charting API such as that demonstrated on www.publicprofiler.org (see Fig. 1). Furthermore, it might be possible to semi-automate the textual description or graphic montages that are driven by the underlying characteristics of the data. As in previous public consultation activities (Longley & Singleton, in press), public feedback on the accuracy of the cluster assignment and descriptions would also be requested.

From a GIS visualisation perspective the advent of Google Maps has raised user expectations in terms of the level of interactivity offered in Web GIS (Haklay et al., in press). However, where cartographic representations are built to a user specification, the technical implementation of a Google Map style Web GIS is a significant challenge (Gibin et al., 2008). Using the Google API as presently configured, display of area data on the fly requires a large amount of processor usage, and as such is not suitable for websites where the re-computation of data coverages is likely to occur frequently. In order to circumnavigate these issues it is possible to use pre-rendered data, such as those created for Maptube (www.maptube.org), however, leads to massive storage requirements, and would not be suitable for an application where coverages are required to be generated for each user request. However, again, like the processing of classifications from raw data, these hurdles may be overcome by deploying solutions across parallel machines with multiple processors.

Consolidation

There has been a sea change in the way in which we conceive of the organisation of social systems. Batty (2008) challenges the view of social science that systems are in equilibrium, and argues that they should be studied from the bottom up, through methods 'that no longer takes the structures that we observe for granted but poses the much deeper question of how they develop'. This view is fundamentally at odds with the old model by which geodemographics are constructed by a small number of individuals from the top down, leading to systems that are deemed to present a 'privileged view' of reality. The development of geodemographics has been valuable in that it has encouraged focus upon the development and use of rich datasets at the local level, yet as time has gone on they have retained the perspective that the needs of real communities could be viewed through the lens of a single general purpose classification.

For a while, the illusion that this remained the best outcome was maintained because the leading commercial classifications had the best data for commercial applications (and arguably this remains the case – specifically not only with regard to income data but also with regard to elements of data infrastructure that the public/academic sectors should better organise). But, more generally, such a perspective is necessarily limited for the simple reason that the best classification for predicting, say, the type of holiday you will buy is not appropriate to understanding your attitudes to public health, community security or private education. This paper has suggested that a new model of geodemographics will be dominated by a problem-centred, domain specific approach led by visualisation, interactive data exploration, and data fusion from a diverse range of sources. There are some indications of this movement already, such as the development of Health ACORN by CACI. From an academic standpoint, geodemographics has always been data led, and this will continue to be the case. There are also implications for public sector data policy and some parts of the public sector need to lighten up about disclosure control, and this needs to be a focus of concerted research effort.

The logical extension to progress in this important domain of applied geography is that the process of creating geodemographic classifications should become more attuned to the processes of social networking, in order to facilitate classification through consultation and the involvement of stakeholders. In this way, the very process of creating classifications should serve as a means of persuasion of both the merits and validity of the exercise. This links concepts to measurement, which lies at the heart of academic critiques of the approach.

References

Anselin, L. (1995). Local indicators of spatial association – LISA. *Geographical Analysis, 27*, 93–115.
Arthur, C., & Cross, M. (2006). *Give us back our crown jewels*. London: The Guardian.
Ashby, D. I. (2005). Policing neighbourhoods: exploring the geographies of crime, policing and performance assessment. *Policing & Society, 15*, 413.
Ashby, D. I., & Longley, P. A. (2005). Geocomputation, geodemographics and resource allocation for local policing. *Transactions in GIS, 9*, 53.
Atkinson, R., & Bridge, G. (2005). *Gentrification in a global context: The new urban colonialism*. London: Routledge.
Bassett, K., & Short, J. (1980). *Housing and residential structure: Alternative approaches*. London: Routledge.

Batey, P., & Brown, P. J. B. (2007). The spatial targeting of urban policy initiatives: a geodemographic assessment tool. *Environment and Planning A, 39*, 2774–2793.
Batey, P., Brown, P. J. B., & Corver, M. (1999). Participation in higher education: a geodemographic perspective on the potential for further expansion in student numbers. *Journal of Geographical Systems, 1*, 277.
Batty, M. (2008). Generative social science: a challenge. *Environment and Planning B, 35*, 191–194.
Birkin, M., Clarke, G., & Clarke, M. (2002). *Retail geography and intelligent network planning*. Chichester: Wiley.
Blake, M., & Openshaw, S. (1994). *GB profiles: A user guide*. Leeds: School of Geography, University of Leeds.
Bowker, G. C., & Star, L. S. (1999). *Sorting things out: Classification and its consequences*. Massachusetts: MIT Press.
Brimicombe, A. J. (2007). A dual approach to cluster discovery in point event data sets. *Computers, Environment and Urban Systems, 31*, 4–18.
Brown, P. J. B., & Batey, P. (1994). *Design and construction of a geodemographic targeting system: Super profiles 1994. WP-40*. Liverpool: Urban Research and Policy Evaluation Regional Research Laboratory, Department of Civic Design, University of Liverpool.
Brunsdon, C. (2006). *A cluster-based approach to the zoning problem using an extended genetic algorithm. GIS research*. UK, Nottingham: University of Nottingham.
Burrows, R., & Gane, N. (2006). Geodemographics, software and class. *Sociology, 40*, 793–812.
Butler, T., Hamnett, C., Ramsden, M., & Webber, R. (2007). The best, the worst and the average: secondary school choice and education performance in East London. *Journal of Education Policy, 22*, 7–29.
CACI. (2005). *Acorn: The smarter consumer classification user guide*. London: CACI.
Castells, M. (1973). *Social theory and the urban question*. London: Arnold.
CLG. (2007). Indices of deprivation 2007. *Government, C.A. L.* London: CLG.
Debenham, J. (2001). *Understanding geodemographic classification: Creating the building blocks for an extension*. Working Paper. Department of Geography, Leeds University.
Debenham, J., Clarke, G., & Stillwell, J. (2003). Extending geodemographic classification: a new regional prototype. *Environment and Planning A, 35*, 1025–1050.
Everitt, T. B. (1974). *Cluster analysis*. London: Heinemann Educational Books Ltd.
Farr, M., & Evans, A. (2005). Identifying 'unknown diabetics' using geodemographics and social marketing. *Interactive Marketing, 7*, 47–58.
Feng, Z., & Flowerdew, R. (1998). Fuzzy geodemographics: a contribution from fuzzy clustering methods. In S. Carver (Ed.), *Innovations in GIS 5*. Taylor & Francis Ltd.
Foster, I., & Kesselman, C. (1999). *The grid: Blueprint for a new computing infrastructure*. San Francisco: Morgan Kaufmann.
Fotheringham, A. S. (1997). Trends in quantitative methods I: stressing the local. *Progress in Human Geography, 21*, 88–96.
Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. Chichester: Wiley.
Gibin, M., Singleton, A., Milton, R., Mateos, P., & Longley, P. (2008). An exploratory cartographic visualisation of London through the Google maps API. *Applied Spatial Analysis and Policy*.
Gordon, A. D. (1999). *Classification* (2nd ed.). Boca Raton, FL: Chapman and Hall/CRC.
Goss, J. (1995). Marketing the new marketing. The strategic discourse of geodemographic information systems. In J. Pickles (Ed.), *Ground truth*. New York: Guildford Press.
Haklay, M., Singleton, A. D., & Parker, C. Web mapping 2.0: the neogeography of the geoweb. *Geography Compass*, in Press.
Harris, R. J., & Longley, P. A. (2002). Creating small area measures of urban deprivation. *Environment and Planning A, 34*, 1073–1093.
Harris, R., Sleight, P., & Webber, R. (2005). *Geodemographics, GIS and neighbourhood targeting*. London: Wiley.
Horlick-Jones, T., Rowe, G., & Walls, J. (2007). Citizen engagement processes as information systems: the role of knowledge and the concept of translation quality. *Public Understanding of Science, 16*, 259–278.
Kempf-Leonard, K. (2005). *Encyclopaedia of social measurement*. San Diego: Elsevier.
Longley, P. A. (2005a). Geographical information systems: a renaissance of geodemographics for public service delivery. *Progress in Human Geography, 29*, 57–63.
Longley, P. A. (2005b). Urban studies. In K. Kempf-Leonard (Ed.), *Encyclopaedia of social measurement*. San Diego, CA: Elsevier.
Longley, P. A., & Goodchild, M. F. (2008). The use of geodemographics to improve public service delivery. In J. Hartley, C. Donaldson, C. Skelcher, & M. Wallace (Eds.), *Managing to improve public services*. Cambridge: Cambridge University Press.
Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2005). *Geographic information systems and science*. John Wiley & Sons.
Longley, P. A. & Singleton, A. D. Classification through consultation: public views of the geography of the E-society. *International Journal of Geographical Information Science*, in press.
Miller, H. J., & Han, J. (Eds.). (2001). *Geographic data mining and knowledge discovery*. London: Taylor and Francis.
O'Reilly, T. (2005). What Is web 2.0 Design Patterns and Business Models for the Next Generation of Software (http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-Web-20.html).
Parker, S., Uprichard, E., & Burrows, R. (2007). Class places and place classes geodemographics and the spatialization of class. *Information, Communication & Society, 10*, 902–921.
Putnam, R. (2000). *Bowling alone: The collapse and revival of American community*. London: Simon & Schuster Ltd.
Rose, D., Pevalin, D., & O'Reilly, K. (2001). *The National statistics socio-economic classification*. London: ONS.
Saunders, P. (1995). *Social theory and the urban question*. New York: Routledge.
See, L., & Openshaw, S. (2001). Fuzzy geodemographics targeting. *Regional science in business*. Berlin: Springer.
Shelton, N., Birkin, M., & Dorling, D. (2006). Where not to live: a geo-demographic classification of mortality for England and Wales, 1981–2000. *Health and Place, 12*.
Shevky, E., & Bell, W. (1955). *Social area analysis: Theory, illustrative application, and computational procedures*. Stanford: Stanford University Press.
Sieber, R. E. Participation, governance and the geoweb. GIScience 2008. Fifth annual conference on geographic information science, 2008, Park City, Utah.
Singleton, A. D. (2007). *A spatio-temporal analysis of access to higher education*. London: Department of Geography, University College London.
Singleton, A. & Longley, P. Creating open source geodemographics – refining a national classification of census output areas for applications in higher education. *Papers in Regional Science*, in press.
Sleight, P. (2004). An introductory review of geodemographic information systems. *Journal of Targeting, Measurement and Analysis for Marketing, 12*, 379–388.
Thrift, N. (2005). *Knowing capitalism*. London: Sage.
Vanderwal, T. (2007). Folksonomy.
Vickers, D., & Rees, P. (2007). Creating the UK national statistics 2001 output area classification. *Journal of the Royal Statistical Society. Series A. Statistics in Society, 170*, 379.
Webber, R. (2004). *The relative power of geodemographics vis a vis person and household level demographic variables as discriminators of consumer behaviour*. CASA Working Paper. London: CASA, UCL.
Webber, R. (2007). The metropolitan habitus: its manifestations, locations, and consumption profiles. *Environment and Planning A, 39*, 182–207.
Weinberger, D. (2007). *Everything is miscellaneous: The power of the new digital disorder*. Baltimore: Henry Holt and Company.
Williamson, T., Ashby, D., & Webber, R. (2006). Classifying neighbourhoods for reassurance policing. *Policing and Society, 16*, 189–218.