6

Spatial representation: the social scientist's perspective

D J MARTIN

This chapter focuses on the representation of socioeconomic phenomena within GIS. Such phenomena include the distribution and characteristics of population, economic activity, and aspects of the built environment. The discussion aims to demonstrate that the association of many such phenomena with precise geographical coordinates is problematic, and there is considerable diversity among the available approaches at both conceptual and technical levels. GIS offer extensive tools for the manipulation and modelling of socioeconomic data, but each option has its own advantages and disadvantages: there are no fundamentally 'right' solutions. There is a need to consider carefully the potential effects of each representation strategy on data output and display.

1 INTRODUCTION

GIS are powerful tools for the manipulation of spatial objects. The application of GIS technology to socioeconomic uses has continued to be a major international growth area, with the population censuses of the 1990s providing an added impetus. Specialised systems for service planning, neighbourhood profiling, and market analysis have been vigorously developed to utilise new digital data (for examples, see Birkin et al, Chapter 51; Longley and Clarke 1995). In many countries, although notably not the United Kingdom, cadastral applications have been some of the major implementations and the utilities are important for GIS users worldwide (see Bibby and Shepherd, Chapter 68). This chapter explores the conceptualisation of socioeconomic phenomena as spatial objects, without assuming any particular application such as marketing or resource allocation. Spatial objects in this sense are entities having both spatial location and spatially independent attribute characteristics (Gatrell 1991). The treatment here of representation as a primarily technical problem permits integration with much other GIS literature, but in so doing it is important to recognise that we do not adequately investigate the very important issues of exactly what is to be represented, who makes the selection, nor the purposes to which it may be put. These issues, particularly relevant to the application of GIS to social phenomena, have been the subject of considerable debate (see for example Pickles, Chapter 4; 1995a), and are also addressed briefly here. All GIS representations are essentially abstract models of selected aspects of the real world, and there will be important questions which cannot be addressed adequately within this framework.

In Chapter 5, Raper considered the discretisation of physical and environmental phenomena for representation in GIS, thereby implementing widely used 'consensus' approaches to conceptualisation. Some phenomena of interest to the social scientist, particularly those relating to the built environment such as property ownership and values, may be treated in much the same way as these physical characteristics. However, the association between many socioeconomic phenomena such as unemployment, population density or ill health and spatial coordinates can be even more problematic than for their physical counterparts. Intuitively, we understand that socioeconomic phenomena vary across geographical space, but their values cannot usually be measured unambiguously at any given location, and it is not clear what (if any) are the fundamental spatial units to which they relate (see also Veregin, Chapter 12).

The rest of this chapter is divided into four sections. The following section reviews the range of potentially spatial socioeconomic data, and considers in more detail the ways in which these may be georeferenced, that is, associated with specific locations. Section 3 extends this discussion to examine the ways in which the basic spatial objects representing socioeconomic phenomena may be represented within a GIS: representation strategies may be divided into those which actually transform the data in order to create different types of spatial object, and those primarily concerned with display. Section 4 then addresses some further considerations relevant to spatial representation in this context, but which do not relate directly to technical processes internal to GIS.

2 GEOREFERENCING SOCIOECONOMIC PHENOMENA

This section considers the links between the enormous variety of socioeconomic phenomena and the relatively limited range of objects which can be used for their spatial representation in computer databases. One of the most basic socioeconomic tasks is the measurement of population. Rhind (1991) divides the sources of large-scale population data into three groups: the conduct of censuses, the maintenance of population registers, and the estimation of population size by indirect means such as the interpretation of remotely-sensed imagery (see also Smith and Rhind, Chapter 47). Additional data sources providing detailed characterisation of the population and its activities include the wide range of social data collected and published by government departments and statistical organisations (statistics pertaining to unemployment, health etc.), and the ever-growing range of information about individuals maintained by commercial organisations as part of their business activities. These include data about flows of many types, such as commuters, migrants, freight, and information. Initially, most of these data do not contain explicit geographical locations in the form of map coordinates, but are associated with the addresses, place names, or regions used in the organisation of service delivery or political representation. In developing GIS applications which attempt to deal with these population-related phenomena, it is thus always necessary to use some form of indirect spatial referencing, frequently via the geographies of the administrative or built environments. In these applications, the GIS user is almost always using secondary data, and should be particularly conscious of data quality and fitness-for-purpose. These socioeconomic data, as collected, can be georeferenced by one of three types of spatial object: areas, lines, or points. Line referencing occurs when only a street or route location is given, providing referencing to a linear object or flow of some kind, but this is actually quite uncommon. The endpoints of such lines are generally points such as street intersections or areal units such as local government areas, and we shall therefore focus on the use of points and areas, in theory and in practice.

2.1 Theory

Population data are most commonly related to geographical locations by reference to areal geographies such as census zones, electoral constituencies, local government areas, or regular grid squares. These are frequently the only areal units for which socioeconomic data are reported, and information derived initially from individuals (either as a sample, or from the whole population), is therefore aggregated to provide summary values for each areal unit. The difficulty with these areal units as a method for georeferencing is that they are essentially ‘imposed’ rather than ‘natural’ units (Unwin 1981). This means that the locations of boundaries may be arbitrarily related to the phenomena which are being measured. This has two aspects: first, a large region may be subdivided into smaller areas at many different scales. For example, the United Kingdom may be divided into around 70 counties, 460 districts, or 10 000 wards. Second, at a given scale, there are different ways of configuring the boundaries of these areal units, each of which results in a different aggregation of the individual-level data. Each of the last three censuses in the United Kingdom has had a different configuration of approximately 10 000 wards. Each reconfiguration would produce a different distribution of zone characteristics, even if there were no change in the underlying population. This is known as the *modifiable areal unit problem* (MAUP), and is discussed more fully by Openshaw and Taylor (1981) and Openshaw (1984). Its effects on spatial analysis have been the subject of continued debate (Fotheringham and Wong 1991), and awareness of the problem is of particular importance when designing zonal systems, as discussed by Openshaw and Alvanides (Chapter 18). The use of areal units with irregular boundaries also presents difficulties for the representation of socioeconomic phenomena on a map, as large, sparsely populated areas will tend to dominate the visual image (but see Elshaw Thrall and Thrall, Chapter 23, for some different cartographic representations). Most zoning schemes of this kind are designed to cover the entire land surface and therefore include extensive areas of unpopulated land, leading to wide variations in population density between areas. The cities in which most people live are represented by small zones covering a very small proportion of the mapped area. A further, and related, difficulty with all types of aggregate data is known as the *ecological fallacy* (Blalock 1964): relationships between variables which are observed at one level of aggregation (e.g. a correlation between household income and educational achievement at county level) may not hold at the individual, or any other, level of aggregation.

An alternative way of thinking about population is as data relating to points. The difficulty here is that it is hard to determine precisely the location of an individual or household. This is usually performed by reference to the home address, for which a location may be derived from a street segment, postal code, or individual property location. Data of this type are most commonly derived from consumer surveys, customer (patient, visitor, etc.) lists, and other address-based lists, including the population registers common in some European countries (Redfern 1989; Ottoson and Rystedt 1991). Individual-level data are hard to visualise in GIS, and many concepts such as an unemployment 'rate' cannot be measured for an individual (who will be registered employed or not), but only have meaning in relation to aggregate data. For some purposes (e.g. employment mapping), the home address may not be the most appropriate spatial reference, but is usually the only option. A full database of individual locations with associated social characteristics would however offer an ideal basis for purpose-specific areal aggregation in which the size and shape of areal units could be redesigned as required.

The modifiable areal unit problem and ecological fallacy, together with the visual constraints associated with zone and point mapping are not restricted to GIS, but also apply to traditional cartographic methods using these types of data. The additional significance of these difficulties in a GIS context is that their effects may be propagated in complex ways through many subsequent operations in which the initial data are transformed for analysis or visualisation, making it difficult to unravel their impact on the final output from the system.

2.2 Practice

In addition to these theoretical considerations, a brief survey of georeferencing in practice reveals the enormous increase in the resolution and range of products available for attaching socioeconomic data to specific locations, illustrated for the case of the United Kingdom in Figure 1. Early census mapping was dependent on rather large and poorly defined spatial units, but over time there has been a massive improvement in the resolution and quality of such data sources. The dual independent map encoding (DIME) system was developed for the 1970 US census, and included records for each street segment, with grid references for street intersections and information about the blocks falling on each side of the street segments. The DIME data structure was an important development (Peucker and Chrisman 1975), but the database covered only major metropolitan areas, and was fundamentally tied to the geography of the built environment (Barr 1996). The original motivation for such databases in the USA was the actual organisation of the census enumeration process, whereas in the United Kingdom mapping products were developed primarily for data display and analysis. For the 1971 UK census, no digital boundary data were available, but census data were published for 1-km grid squares nationally and 100-m squares in urban areas, providing a direct method for the production of maps such as those in the census atlas *People in Britain* (CRU/OPCS/GRO(S) 1980). A single (subjectively determined) centroid location was also provided for each enumeration district (ED) – the smallest areal units for which census data are published. By the early 1980s DIME had been extended to cover many more areas and in the UK a set of digital boundaries was produced at the ward level (each ward typically contains around 13 EDs) in addition to the ED centroids. With the massive growth in GIS in the mid 1980s, a number of organisations began to produce additional digital data for socioeconomic zones in local areas, such as the production of ED boundaries. In the UK, where there is no direct correspondence between census and postal geographies, there was also a marked increase in interest in postal geography as a georeferencing system (Raper et al 1992), and a national directory called the Central Postcode Directory (CPD) came into widespread use, containing a 100-m grid reference for each unit postcode in the country, typically covering only 15 addresses, compared to 169 in the average ED.

The early 1990s have seen another major increase in the completeness and resolution of basic georeferencing products. In the USA, the Topologically Integrated Geographic Encoding and Referencing (TIGER) system now provides national coverage of DIME-type information, with extensive topology, information about the shape of street segments and relationships with a variety of other statistical and administrative zoning systems (Broome and Meixler 1990). The TIGER database has been maintained and made widely available, ensuring continual enhancement, as data suppliers take the basic files and 'add value' by integration with other datasets. In the United Kingdom, two commercially produced digital boundary sets were produced for 1991 census EDs, together with a directory indicating the intersections between EDs and unit postcodes. 1996 saw the completion of ADDRESS-POINT, a national database containing 0.1-m grid references and unique property reference numbers (UPRNs) for all properties, which will provide the basis for further derived products in the future (Smith and Rhind, Chapter 47; Martin and Higgs 1996). Similar enhancements to the resolution of georeferencing products have occurred elsewhere, with an increasing focus on individual addresses (e.g. Lind and Christensen 1996). It should be noted, however, that the process of associating socioeconomic data with appropriate georeferences is frequently error-prone, because of the inherent uncertainty involved in the integration of datasets created by different organisations at different times, yet purporting to describe the same entities. Recent initiatives to develop national standards for address referencing in the UK, for example, are still a long way from widespread implementation (Cushnie 1994).

The preceding discussion illustrates that there have been increases in both the number of products available, and in spatial resolution over the last two decades. It is tempting to assume that georeferencing is all that is involved in the representation of socioeconomic phenomena, yet despite the improvements noted above, none of these products provides fundamentally the 'correct' spatial object for representation in GIS. As explored in the following section, many alternative representations of these same data sources may be constructed using the manipulation tools available within GIS software. This can be done either by transforming the input data to another type of spatial object, such as a continuous surface model, or by transforming geographical space for the purposes of visualisation, for example in cartogram construction.

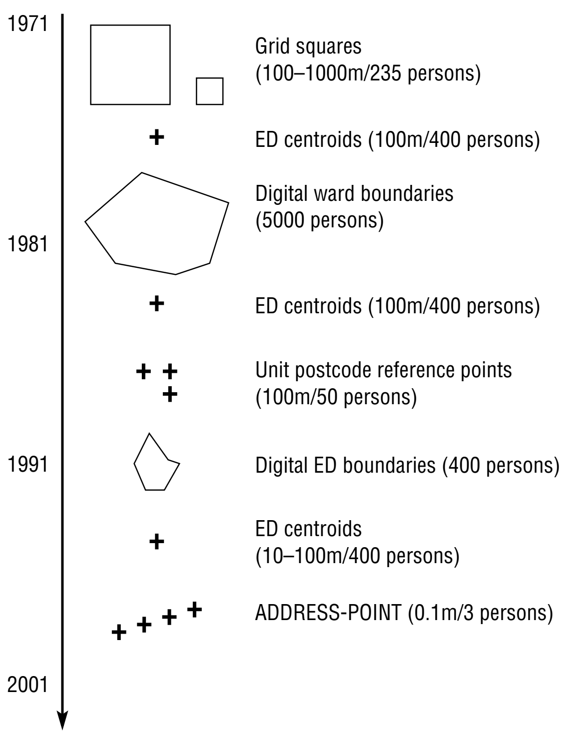

Fig 1. Increases in georeferencing resolution in the UK, 1971–present.

3 SPATIAL REPRESENTATION OF SOCIOECONOMIC PHENOMENA

The previous section has provided a discussion of the spatially-referenced data available to describe various phenomena of interest to the social scientist. Although there may be many thousands of data series in existence which describe aspects of the socioeconomic world, there are relatively few options available for actually attaching these measurements to specific geographical locations. It is important to understand the relationship between these geographical data series, usually produced externally to the GIS, and the different representation strategies which may be adopted within a GIS.

This discussion is best introduced by a simple example: to attach disease incidence to household addresses and thus to point grid references assigns a particular type of representational model to that disease. This will tend to promote a different way of thinking about the disease to that which might have been adopted if cases were treated as flows, assigned to census zones or modelled as some kind of continuously varying surface phenomenon or field (Figure 2). Different analytical techniques may be used on the different representational models, and different answers may result from questions concerning the degree of clustering in disease incidence, or its association with the geographical distribution of some other variable (see also Gatrell and Senior, Chapter 66). In the following chapter, Mark (Chapter 7) discusses the issue of deciding what is to be represented in more detail, making the distinction between 'entities' which exist in the real world and 'objects' which are part of the digital representation. An important feature of GIS is the ability to remodel data from one spatial object type to another, including the generation of complex objects from simple or primitive ones, as identified for example by Gatrell (1991; see also Fisher, Chapter 13).

It is useful to consider representation not as a single operation, but as a process, whereby selected

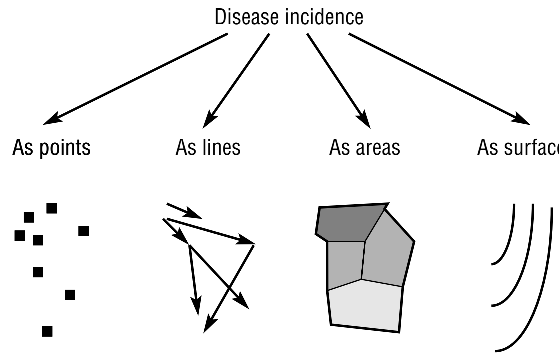

Fig 2. Different spatial conceptualisations of disease incidence.

information about a real-world entity passes through a number of stages along the road to visualisation or some other form of data output, each of which may affect the way in which its spatial characteristics are stored and accessed. This discussion is organised around the examples given in Table 1. Much has been made in GIS literature of the distinction between vector and raster data structures (e.g. Maffini 1987; see also Batty, Chapter 21). Although important to the operation of the GIS at a technical level, these issues are largely independent of the issues being considered here, and the different approaches to representation may be implemented with relative ease using each of the available data structures.

The columns of Table 1 illustrate four 'classes' of geographical phenomena, namely points, lines, areas, and surfaces. As will be seen, this division based on spatial dimensionality is not without its difficulties, but it will serve here to illustrate some of the many

Table 1 Examples of the representation of socioeconomic phenomena in GIS.

	Point	*Line*	*Area*	*Surface*
Real-world entity	individual disease case?	journey to work	property ownership	population density?
Digital object	property or postcode coordinate	street segment coordinates	land parcel, census zone boundary	TIN or elevation model
Manipulation techniques	nearest neighbour analysis, boundary generation, surface generation	network functions, topological analysis	areal interpolation, centroid generation, surface generation	slope analysis, TIN creation, DEM creation, analysis of surface form
Visualisation techniques	point mapping, multivariate glyphs, convert to 3D	line mapping, line cartograms, convert to 3D	choropleth mapping, area cartograms, convert to 3D	isoline mapping, TIN mapping, grid mapping, convert to 3D

possible approaches to representation. The rows of the table illustrate four stages in the representation process, which are further illustrated in Figure 3. Geographical entities are considered to exist in the real world; georeferencing provides the link with digital objects which may be used to represent the locations of such entities; GIS provide manipulation tools for the creation of new objects, moving both within and between object classes; and finally, visualisation techniques may be applied in each case. Between these stages are: (1) data collection and entry, (2) data manipulation, (3) data output transformations discussed in Martin (1996), which are analogous to the transformation stages in the traditional cartographic process described by Clarke (1995). In traditional cartography the printed map embodies a single representation of each real-world entity, but in GIS there is a variety of representation strategies available. Many forms of GIS analysis may not actually result in visualisation, but will produce some other kind of non-graphical output (query results, statistical summaries, etc.) directly by manipulation of the digital objects. Another important parallel with cartography is that the relationship between the 'real' world and representation at each stage is not simply determined by the accuracy of coordinates and attributes, but is influenced by operator decisions about which phenomena are to be included, and how they are to be measured, classified, and symbolised.

In the first row of Table 1 are examples of different real-world entities which are of interest to the social scientist. For some such entities there is general consensus regarding their object class, for example a journey to work is conceptualised as a linear feature or flow, and legal land ownership by definition relates to a specified parcel of land, the ownership documents often being accompanied by some form of map or description of the area enclosed. The other two examples are more ambiguous: as in the example above, most people would probably think of disease incidence in terms of a pattern of individual points over space, but others might argue that these too are more correctly considered as flows or even as a continuously varying surface. Population density is tentatively included as an example of a real-world entity which may actually be a surface, although this is again subject to some debate and is further addressed below. In many cases, the 'true' class depends on the scale at which the object is being considered, thus population density which at very large mapping scales may be considered as a point pattern is more convincingly a surface at smaller scales.

The second row of the table contains examples of digital objects falling into each of the classes. These are the kinds of spatial 'data' which have been discussed above, and which are used for the georeferencing of many socioeconomic phenomena which do not have any precisely measurable locational characteristics. The coordinates of a property will frequently be used as a proxy for the location of an individual member of the population, perhaps with a high degree of spatial precision, or perhaps less precisely via the grid reference of the corresponding postcode. Encoding of a journey to work, or other route information, is most commonly accomplished by reference to line segments in the spatial database or perhaps to a computed line

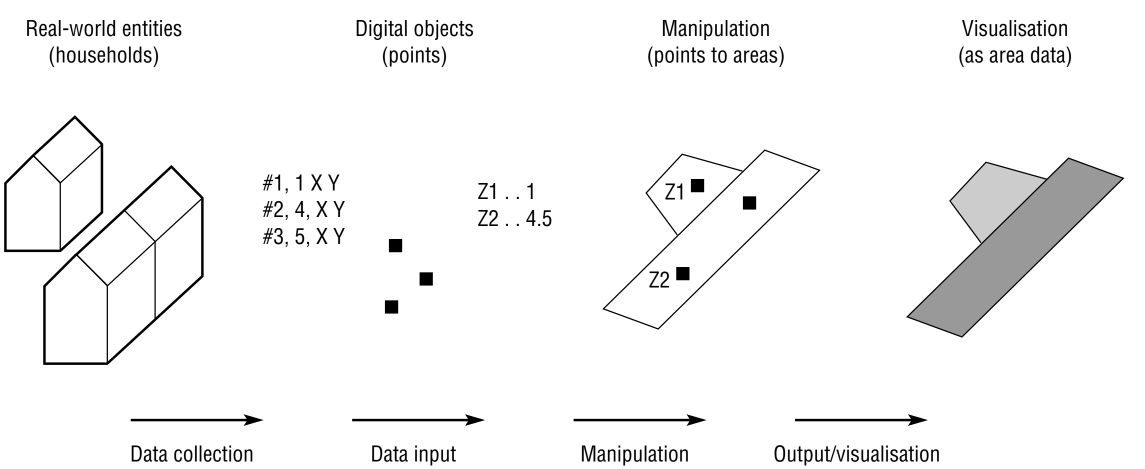

Fig 3. Representation in GIS viewed as a process.

between two address locations. In this context, it is appropriate to mention non-Cartesian concepts of space, in which relations such as time or cost may replace Euclidean distance in defining the separation between locations (see Couclelis, Chapter 2). Such spaces may indeed be the most appropriate frameworks within which to explore certain socioeconomic phenomena, but they are not easily handled by contemporary GIS data structures, and there has been little development work in this area. Areal phenomena are readily encoded by a series of coordinates defining an enclosed area. It should be noted, however, that the use of areal objects to describe such phenomena as population results in a change of object class from point or surface (depending on the user's point of view) to area. It is not possible to record directly any social phenomena in surface form, although various representation strategies exist for the storage of surface models – some of which are discussed below.

The third row of Table 1 contains examples of a range of manipulation functions which may operate on spatial objects within a GIS. These include both techniques for the analysis of spatial form and tools for the transformation of objects from one class to another. Spatial analysis techniques are dealt with more fully in later chapters in this volume by Openshaw and Alvanides (Chapter 18), Getis (Chapter 16), and Fischer (Chapter 19). Spatial analytical tools are heavily influenced by the object class of the models on which they operate, and the decision to represent phenomena using different models will therefore have a significant impact on the types of analysis which may be performed (Bailey and Gatrell 1995). The transformation of objects between classes includes, for example, the generation of boundaries around point data by the creation of Thiessen polygons (Boots, Chapter 36; 1986). This allows both data associated with the original points to be mapped as areal data, and areal manipulation and analysis functions to be applied. An extensive series of functions exists for the analysis of network topology and routing problems, which work on suitably structured line data. Although precise distance calculations are possible along such networks, it is frequently difficult to define the endpoints of flows or journeys with sufficient detail, and single origin and destination points are frequently used to represent the flows between zones. The definition of such centroids is another transformation operation, permitting values originally relating to zones to be applied to points and thus mapped and analysed as point patterns. The acceptability of the assumptions involved in each of these types of estimation will vary according to the specific application. A frequently encountered problem with socioeconomic data is the need to relate phenomena recorded for two incompatible sets of areal units. A variety of approaches has been suggested, including the use of ancillary variables to aid in the interpolation of data values between the two sets of boundaries, and the intermediate estimation of values in an underlying surface (Flowerdew and Green 1991; Goodchild et al 1993).

This last approach indicates that it is also possible to remodel data from each of the other source object classes into surface form. The common practice of georeferencing population-related phenomena by reference to discrete locations such as address points or census zones has tended to reinforce the view that these phenomena are indeed discrete, but there are good grounds for reconceptualising many such phenomena as continuously varying over space, and attempting to represent them in this way. Nordbeck and Rystedt (1970) describe phenomena such as population density as 'reference interval functions' which cannot be measured at a single point, but only have meaning in relation to some reference interval, such as 200 persons per hectare. Such functions may be treated as surfaces with validity. None of the existing data models used by GIS is able to represent fully continuous variation over space, but structures such as altitude matrices and Triangulated Irregular Network (TIN) models can represent surface form adequately if used at an appropriate scale. Tobler (1979) presents a volume-preserving approach to the construction of socioeconomic surfaces from areal data, and Martin (1989) illustrates surface construction from zone centroids, a technique developed further by Bracken and Martin (1995), for example.

The final row of the table illustrates the kinds of display and visualisation technique which may be applied to spatial data representing socioeconomic phenomena. Visualisation may involve the direct display of data values using a visual representation of the same object class, or may involve further transformation of the data for display which does not directly affect the form of the digital objects in the database. Traditional cartographic representation makes use of points, lines, and areas together with particular conventions for symbolisation, projection, and scale (Elshaw Thrall and Thrall, Chapter 23; Kraak, Chapter 11). These techniques may be directly reproduced using GIS technology, but a broader range of possibilities also becomes available. Gatrell (1994) explores the visualisation of point patterns, which may be applied to the types of postal and address information discussed above, but the use of choropleth (shaded area) maps for areal data is still the most widespread type of socioeconomic data mapping, embodying all the difficulties associated with modifiable areal units. Flow data are conventionally represented by mapped lines, while surfaces are variously depicted by isolines (lines of equal value) such as contours, sample points, or quasi-continuous shading of small areas – such as that which may be output from an elevation model comprising a grid of small cells. Less conventional, but potentially more powerful, representations of each object class may be achieved by the use of mapping tools in which strict geographical relationships are relaxed in order to represent other features of the data. Dorling (1994) presents a range of such approaches, including multidimensional glyphs for point data and area cartograms in which the geographical area of zones is replaced by a non-spatial attribute such as population size in determining the size of the mapped symbol.

There is increasing interest in 3-dimensional visualisation tools in GIS, both in order to recreate 'realistic' scenes and as an aid to the understanding of patterns in data. In architectural and planning applications, there is increasing integration between GIS and 3-dimensional tools for the exploration of urban scenes which may be affected by redevelopment, for example (Levy 1993; Liggett and Jepson 1995). Data visualisation in three dimensions is less well developed in socioeconomic applications, but there is considerable scope for the use of 3-dimensional scenes in which some of the visual parameters – such as the colour and texture of objects – are used to symbolise otherwise unobservable pattern in socioeconomic phenomena. An interesting example is provided by Wood et al (1996), in which 3-dimensional models are used to explore features of the population structure of Greater London.

4 DISCUSSION

In the light of these many different possibilities, is there a single 'right' way to represent socioeconomic phenomena within GIS? The answer to this question is almost certainly 'no', the best strategy being highly dependent on the specific application. The implication of this is that considerable onus is placed on users to fully understand the implications of the representation strategy which they choose to adopt. In addition to the technical and conceptual considerations already addressed, a number of further relevant issues are addressed here, although fuller discussions of some of these will be found in later chapters. (e.g. see Cova, Chapter 60, for a discussion of representational issues in emergency management; Larsen, Chapter 71, for issues in environmental monitoring and assessment applications).

In concentrating on spatial representation, this discussion has largely assumed that the spatially independent attribute characteristics of objects are unproblematic, but there will be many situations in which this is not the case. Attribute values will be subject to broadly similar operator selectivity and measurement error to their spatial counterparts, and the extent of these influences may vary over space. Data input to GIS, and therefore the range of issues which can reasonably be addressed, is usually constrained by the questions which were asked in some previously existing survey or census (see Goodchild and Longley, Chapter 40). The 1991 UK Census was subject to differing degrees of underenumeration, which is generally understood to have been highest among the young adult male populations concentrated in large metropolitan centres. In the analysis of the resulting data, there may therefore be complex interactions between the spatial and attribute characteristics. The use of GIS to optimise zoning schemes for socioeconomic data, as illustrated by Openshaw and Alvanides (Chapter 18) and by Openshaw and Rao (1995), potentially offers important advances in the control of attribute variation by spatial manipulation.

Other GIS developments which potentially offer benefits to the representation of socioeconomic phenomena include the incorporation of fuzzy concepts, and non-Euclidean spaces (see Couclelis, Chapter 2, and Fisher, Chapter 13). Openshaw (1989) uses the term 'fuzzy geodemographics' to refer to an approach in which the inherent uncertainty in spatial referencing and attribute values for socioeconomic GIS data are acknowledged. By contrast, in some current applications, where all the information to be processed relates to the same set of zones, or the task is essentially one of list matching (for example, between entries in a directory of post/zip codes), then it may not be necessary to use GIS at all (Barr 1993).

A longstanding issue when dealing with information concerning individuals in computer-readable form has been that of confidentiality, and the need to protect individuals from the inadvertent disclosure of personal information. This has usually been a guiding principle in the publication of data from censuses by government, typically resulting in the imposition of a population threshold size below which no data are published. However, this deliberate aggregation of data runs counter to many business applications, in which an objective is to identify individuals or households with particular characteristics as precisely as possible, and there is considerable commercial interest in the association of different data series in order to build up detailed individual-level information. The 'representation' of individuals in this way raises ethical issues which have not really been addressed by most GIS users, but which are considered more fully by Curry (Chapter 55). The use of GIS in this area has been the subject of broad-ranging critique, such as that found in Pickles (Chapter 4; 1995b). There is not space here to develop a response to these arguments, but it is certainly true that maps have always been tools associated with the exercise of power, representing both the physical and socioeconomic worlds selectively, for the purposes of particular groups and individuals (Wood 1992). These possibilities are multiplied rather than reduced by GIS, with its multiple representational strategies. There is still a considerable gap between the academic critics and proponents of GIS technology, but at present many of these issues would probably not even be recognised by many non-academic GIS users who routinely handle socioeconomic data. As with the technical choices concerning representation, there is a heavy responsibility placed on GIS users to be fully aware of the implications of their actions.

5 CONCLUSION

This chapter has considered the representation of phenomena likely to be of interest to social scientists within GIS. GIS offer a wide range of tools for the manipulation and analysis of models of geographical reality stored as digital objects, but the primary difficulty in socioeconomic applications concerns the most appropriate way in which to measure and model the phenomena of interest. There are many sources of computer-readable data about population and economic activity, but these are not usually associated with directly measurable geographical locations, and some form of indirect spatial referencing is required. The association of population-related phenomena with discrete geographical locations poses technical, conceptual, and ethical difficulties. There have been improvements in the number and resolution of available datasets for georeferencing, but this has not made the conceptualisation of phenomena such as unemployment or population density any easier. There remains disagreement over the precise spatial form of these phenomena, and therefore the most appropriate technical solutions which should be adopted. Initial decisions about georeferencing affect all subsequent transformations of the data, both for the creation of new digital objects and for visualisation and output. GIS users in this area also need to address more carefully some of the broader questions about representation which have perhaps to some degree been provoked by uncritical use of the available data sources.

References

Bailey T C, Gatrell A C 1995 *Interactive spatial data analysis*. Harlow, Longman/New York, John Wiley & Sons Inc.

Barr R 1993 Mapping and spatial analysis. In Dale A, Marsh C (eds) *The 1991 Census user's guide*. London, HMSO: 248–68

Barr R 1996 A comparison of aspects of the US and UK censuses of population. *Transactions in GIS* 1: 49–60

Blalock M 1964 *Causal inferences in nonexperimental research*. Chapel Hill, University of North Carolina

Boots B N 1986 *Voroni (Thiessen) polygons* Concepts and Techniques in Modern Geography 45. Norwich, Geo-Books

Bracken I, Martin D 1995 Linkage of the 1981 and 1991 censuses using surface modelling concepts. *Environment and Planning A* 27: 379–90

Broome F R, Meixler D B 1990 The TIGER database structure. In Marx R W (ed.) *The Census Bureau's TIGER system*. Bethesda, ACSM: 39–47

Clarke K C 1995 *Analytic and computer cartography*, 2nd edition. Englewood Cliffs, Prentice-Hall

CRU/OPCS/GRO(S) 1980 *People in Britain: a census atlas.* London, Her Majesty's Stationery Office

Cushnie J 1994 A British Standard is published. *Mapping Awareness* 8: 40–43

Dorling D 1994a Cartograms for visualising human geography. In Hearnshaw H M, Unwin D J (eds) *Visualisation in geographical information systems*. Chichester, John Wiley & Sons: 85–102

Flowerdew R, Green M 1991 Data integration: statistical methods for transferring data between zonal systems. In Masser I, Blakemore M (eds) *Handling geographic information: methodology and potential applications*. Harlow, Longman/New York, John Wiley & Sons Inc.: 38–54

Fotheringham A S, Wong D W S 1991 The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A* 23: 1025–34

Gatrell A C 1991 Concepts of space and geographical data. In Maguire D J, Goodchild M F, Rhind D W (eds) *Geographical information systems: principles and applications*. Harlow, Longman/New York, John Wiley & Sons Inc. Vol. 1: 119–34

Gatrell A C 1994 Density estimation and the visualisation of point patterns. In Hearnshaw H M, Unwin D J (eds) *Visualisation in geographical information systems*. Chichester, John Wiley & Sons: 65–75

Goodchild M F, Anselin L, Deichmann U 1993 A framework for the areal interpolation of socioeconomic data. *Environment and Planning A* 25: 383–97

Levy R M 1995 Visualisation of urban alternatives. *Environment and Planning B: Planning and Design* 22: 343–58

Liggett R S, Jepson W H 1995 An integrated environment for urban simulation. *Environment and Planning B: Planning and Design* 22: 291–302

Lind M, Christensen T 1996 The address as key to GIS. *Proceedings of the second joint European conference and exhibition on geographical information*. *Barcelona* Vol. 2: 868–77

Longley P, Clarke G 1995a *GIS for business and service planning.* Cambridge (UK), GeoInformation International

Maffini G 1987 Raster versus vector data encoding and handling: a commentary. *Photogrammetric Engineering and Remote Sensing* 53: 1397–8

Martin D 1989 Mapping population data from zone centroid locations. *Transactions of the Institute of British Geographers* NS 14: 90–7

Martin D 1996 *Geographic information systems: socioeconomic applications*, 2nd edition. London, Routledge

Martin D, Higgs G 1996 Georeferencing people and places: a comparison of detailed datasets. In Parker D (ed.) *Innovations in GIS 3*. London, Taylor and Francis: 37–47

Nordbeck S, Rystedt B 1970 Isarithmic maps and the continuity of reference interval functions. *Geografiska Annaler* 52B: 92–123

Openshaw S 1984 *The modifiable areal unit problem.* Concepts and Techniques in Modern Geography 38. Norwich, Geo-Books

Openshaw S 1989b Learning to live with error in spatial databases. In Goodchild M F, Gopal S (eds) *Accuracy of spatial databases.* London, Taylor and Francis: 263–76

Openshaw S, Rao L 1995 Algorithms for reengineering 1991 Census geography. *Environment and Planning A* 27: 425–46

Openshaw S, Taylor P J 1981 The modifiable areal unit problem. In Wrigley N, Bennett R J (eds) *Quantitative geography: a British view.* London, Routledge: 60–70

Ottoson L, Rystedt B 1991 National GIS programmes in Sweden. In Maguire D J, Goodchild M F, Rhind D W (eds) *Geographical information systems: principles and applications*. Harlow, Longman/New York, John Wiley & Sons Inc.Vol. 2: 39–46

Peucker T K, Chrisman N 1975 Cartographic data structures. *The American Cartographer* 2: 55–69

Pickles J (ed.) 1995a *Ground truth: the social implications of geographic information systems*. London, Guilford Press

Pickles J 1995b Representations in an electronic age: geography GIS and democracy. In Pickles J (ed.) *Ground truth: the social implications of geographic information systems*. London, Guilford Press: 1–30

Raper J F, Rhind D W, Shepherd J W 1992 *Postcodes: the new geography*. Harlow, Longman

Redfern P 1989 Population registers: some administrative and statistical pros and cons. *Journal of the Royal Statistical Society* A 152: 1–28

Rhind D W 1991 Counting the people. In Maguire D J, Goodchild M F, Rhind D W (eds) *Geographical information systems: principles and applications*. Harlow, Longman/ New York, John Wiley & Sons Inc. Vol. 2: 127–37

Tobler W R 1979 Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association* 74: 519–36

Unwin D J 1981 *Introductory spatial analysis.* London, Methuen

Wood D 1992 *The power of maps*. London, Routledge/New York, Guilford Press

Wood J, Unwin D, Stynes K, Fisher P, Dykes J 1996 Using the landscape metaphor to understand population data. *http://www.geog.le.ac.uk/argus/poplandscape/index.html*