THE UNIVERSITY OF HULL

Cartographic Information Systems Research Group
C.I.S.R.G.

Hon. Co-ordinator:
Dr. M. Visvalingam
Department of Computer Science
The University
HULL HU6 7RX

Tel. (0482) 465295/465951
Telex 592530

Discussion Paper Series Editors : Dr. M. Visvalingam
Dr. M. E. Turner

C.I.S.R.G. DISCUSSION PAPER 3

Standardising Basic Spatial Units: Problems and Prospects

by

M. Visvalingam

Not to be quoted without the author's permission

September 1988

List of CISRG reports:

Discussion Papers

1. Wade, P., Visvalingam, M. and Kirby, G.H. (1986)
   From Line Geometry to Area Topology, 48 pp.

2. Visvalingam, M. (1987)
   Problems in the Design and Implementation of a GKS-based User Interface for a Graphical Information System, 18 pp.

3. Visvalingam, M. (1988)
   Standardising Basic Spatial Units : Problems and Prospects, 19 pp.

Special Issues

Visvalingam, M. (ed) (1986)
Research based on Ordnance Survey Small-Scales Digital Data
Proceedings of a meeting held as part of the 1986 Annual Conference of the Institute of British Geographers (Reading, Jan 1986), **sponsored by the Ordnance Survey,** 79 pp.

Visvalingam, M. and Kirby, G. H. (1987)
Directory of Research and Development based on Ordnance Survey Small Scales Digital Data, **sponsored by the Ordnance Survey,** 38 pp.

CONTENTS

PREFACE

The multidisciplinary Cartographic Information Systems Research Group (CISRG) of the University of Hull was established in 1985. Following the establishment of interdisciplinary links through the CISRG, the Departments of Computer Science and Geography initiated a third year option in Computer Cartography in 1987. The course is taught by staff in both Geography and Computer Science and is open to final year students in both Departments.

As a supplementary part of this course, students were provided with some background information on controversial topics and were expected to discuss them within the context of a debate. One of these debates focused on the proposals for defining a standard set of **basic spatial units** (BSUs). The choice of BSUs is a critical decision in the construction of all spatial information systems. The aim of the debate was to make students aware of the various factors which constrain the design of BSUs.

This paper arises from the debate and ensuing discussion. It discusses the main topics of interest, namely spatial and aspatial descriptors, the focus of interest, accuracy, shape, homogeneity, modifiability and coverage, size, ecological fallacy, variability of base populations, aggregational flexibility, stability, usage and computing considerations. Some of these issues are relevant to all applications; others relate mainly to the choice of spatial frameworks for collection, distribution and analysis of spatial statistics. I hope that this summary of the properties of BSUs will promote the evaluation of proposals for standard BSUs by a wider community of practitioners.

M. Visvalingam
September 1988

1. INTRODUCTION

The choice of Basic Spatial Units (BSUs) is a critical decision in all spatial information systems, whether manual or computerised. BSUs are the smallest geographic entities for which information is collected and/or made available. Although computer cartography and Geographical Information Systems (GIS) are concerned with the handling of point, line and areal entities, the current discussions on BSUs appear to centre almost exclusively on the need for a standard set of basic **areal units** for cross-referencing various types of personal, socio-economic and other related data.

In Britain, the reporting of personal data is restricted by a traditional respect for their privacy and confidentiality and more recently by the 1984 Data Protection Act. Also, many policy-related and commercially-oriented analyses are concerned with identifying target areas, containing target populations, rather than with identifying individuals or individual households per se. Consequently, data relating to persons and households are released in aggregate form for a set of BSUs which already exist for operational purposes or which are specially designed for purposes of reporting data.

There are at present a variety of such BSUs, which correspond to units of observable phenomena, to functional or administrative units, to measurement units or to units for reporting sensitive data. These various units are seldom spatially coincident and therein lies the source of many problems in handling geographic information.

Academic and commercial users of data, collected by government, are attracted by their proxy value, particularly when used with data from other sources. For example, the 1981 population census Small Area Statistics have been used very extensively in conjunction with market research and/or a firms's own data in locational analyses, marketing, retailing and advertising.

However, data from different sources cannot be compared easily or accurately if they relate to very different sets of BSUs. The move towards standardisation is unconcerned with the use of data units within an organisation. It is mainly concerned with the design of reporting units. It may, however, be possible that the proposed reporting units are smaller than existing operational units. This requires that organisations hold proprietary data in a relatively unaggregated form to satisfy both operational and reporting purposes.

The Chorley Committee on "Handling Geographic Information" (DoE, 1987) made the following recommendations. Geographic data, relating to land areas of the UK, should be referenced directly or indirectly to the National Grid or Irish Grid (Recommendation 34). Data suppliers should both keep and release their data in as unaggregated a form as possible (38). The preferred bases for holding/releasing socio-economic data should be addresses and unit postcodes (39). There should be a consistent scheme for the Grid referencing of addresses and unit postcodes (40).

Even if these proposals are accepted and acted upon, there will be a continuing need for the design of BSUs for different applications. A checklist of various properties of BSUs, which constrain their design, is of help to students of GIS. Some of these properties are listed below and I hope that readers will help me in evolving a more complete and coherent account of these constraining factors. I would also appreciate comments on relative priorities for different applications since the design of all artefacts involves both objective and subjective decisions, judgement and tradeoffs.

2. SCOPE FOR AND CONSTRAINTS UPON DESIGN

The term BSU is consistently used within all applications to denote the spatial primitive. However, the term, spatial primitive, represents conceptually different entities within different classes of applications. This is because different applications operate at different levels within the scenario of GIS and are concerned with different levels of abstraction of spatial reality. In particular, the spatial primitives in one class of application may correspond to complex spatial structures or higher level aggregate units within another class of application. Even within a single application, different spatial frameworks may be appropriate for various purposes, such as measurement, reporting, monitoring, analyses and other operational purposes.

The nature of spatial phenomena is such that it would be unrealistic to advocate a standard set of BSUs for all applications. However, different classes of applications may benefit and progress through adoption and use of standards appropriate to their brief. Let us consider some of the main issues relating to the discussion on BSUs.

2.1 Spatial versus aspatial data

There are two types of data associated with BSUs, namely their spatial definitions and their aspatial or substantive characteristics. Spatial data provide locational and topological information on BSUs. These may take the form of fully structured boundary files or **spatial surrogates,** such as **visual centroids** or grid references of addresses or unit postcodes. Aspatial data describe the identity (through **nominal references** in the form of textual or codified names) and character of the spatial unit. The postal address, the unit postcode and code names are examples of nominal references. The aspatial data, and spatial surrogates when used, are stored in **attribute files.** For very many applications only attribute files need to be in computer readable form.

However, multipurpose corporate information systems need to integrate both spatial and aspatial data. The BSUs must be designed with respect to all proposed applications even if there are no plans to digitise boundaries in the initial phases of a computerised GIS.

Digital boundaries have in the past been excluded from some GIS for reasons of cost. Walter Smith (Mapping Awareness, 1988) and others have suggested that many decision-support systems do not require boundaries. Indeed, the output of many such systems consist of address lists for mailshots and leafletting. However, Visvalingam and Kirby (1984) have argued that visualisation is essential for validation of results in exploratory data analysis. Also, spatial surrogates are insufficient for graphic interaction with GIS. The future will see a growing demand for digital boundaries although the requirements for accuracy may vary with the application.

2.2 Accuracy

The accuracy of spatial boundaries is an important property. Caricatures may be automatically extracted from accurate data for use within small scale applications but graphic enhancement of poor quality data is inadequate for many Land Information Systems. Thus, the long-term national digitising effort must address the requirements for accurate data.

2.3 Focus

A number of decision support systems focus, not on land, but on individuals and/or households. The 1984 Data Protection Act only provides such applications with access to aggregate spatial statistics, such as the population census Small Area Statistics. The primitive for aggregation is the street address; the land parcel corresponding to this address is of little interest to many applications. If street addresses were Grid referenced, data on individual households could be released for any set of spatial units. However, the aggregation of data may be based on non-spatial data. For example, census returns on individual households were linked to the census enumeration district (ED) in past population censuses. Systematic code names for the hierarchies of census reporting units were then used for further aggregation of data.

The term, basic spatial unit, is misleading in this context. Many applications are not concerned with precisely defined boundaries and units of space. Instead, they focus on the analysis of data for basic population units (BPUs). The unit postcode is such a BPU. The precise boundaries of unit postcodes are not known but it is generally accepted that all BPUs, like all BSUs, must be tagged with a unique Grid reference for spatial data analysis.

2.4 Shape

BSUs may be classified into two major types, namely regular and irregular units. Regular units, such as grid squares, are the result of a systematic division of space into areas of the same shape, even if not the same orientation or size. Such systematic divisions of space into regular geometric shapes are called **tessellations** (Peuquet, 1984). Regular tessellations result in units of the same size, while nested tessellations allow the recursive subdivision of a cell into smaller and smaller units. Hybrid tessellations can include more than one type of shape of varying size. All tessellations provide a framework for recording data about arbitrary units of space. The units are arbitrary in the sense that cell boundaries do not correspond to phenomena of interest. Information on the character and distribution of spatial phenomena has to be inferred from the data for these arbitrary units. The widely used grid square framework is one form of tessellation.

Space may also be divided into irregular units. Irregular units are **idiographic** in the sense that each unit may have its own peculiar shape with detached parts and holes. Irregular units may be subclassed into natural, functional and primary units. **Natural units** define the extent of observable phenomena, such as vegetation, soil, geologic and land use types, which are described by nominal or discrete categories of data.

Functional units are specially designed for some specific purpose, such as administration, taxation, ownership, targeting services, policing, for the performance of statutory functions or for the delivery of mail. Functional units are almost always synthetic and given that environmental, population, socio-economic and several other characteristics are all subject to change, functional units are by nature **dynamic.** Unit postcodes and the 1984 Travel to Work Areas are two examples of irregular functional units.

Within a paper-based information system, these natural, legal and functional units will be identified on a map. Owing to the specialised nature of data collection and dissemination and the limitations of a paper-based model for communication of information, data from different sources are held on different map sheets, called **coverages.**

Many applications need to relate information from different sources. This cross-referencing of data is facilitated by the use of consistent names for the same units on the ground. More often than not, the spatial units may not be exactly coincident. This requires the visual overlay of a set of maps so that the characteristics of some specified unit of space may be identified. In a multi-source or corporate GIS, the plots of land which result from the combination of different natural and functional units on different coverages will be assigned unique identifiers, such as the **unique**

property reference numbers (UPRNs) proposed in the Department of Environment's 1972 General Information Systems for Planning (GISP) report and in the Tyne and Wear Area Joint Information System (JIS). This UPRN can then be used to cross reference data through use of a **gazetteer,** which is basically a type of codebook or index to codes (see DoE, 1987, p 166).

The plots of land, to which these UPRNs refer, may be regarded as the **primary units** which make up other user perceived irregular objects. Thus, the latter are intrinsically aggregate, higher-level spatial units and the plot of land is the ultimate spatial primitive. The term **primitive region,** has been used to denote this spatial primitive within information systems (Visvalingam et al, 1986; Kirby et al, 1989). This is because the more frequently used term, **land parcel,** is used in various ways to denote, for example, a unit of land ownership by Her Majesty's Land Registry, a space subject to tax by the Inland Revenue or a unit of land use for the purposes of the Annual Census of Agriculture (Dale, 1988). These various usages of the term land parcel correspond to sets of application-defined spatial units on different coverages.

Digital versions of the Ordnance Survey large scale plans will not in themselves provide full spatial definitions of primitive regions. The data will have to be restructured to yield a complete and coherent set of the land parcels shown on Ordnance Survey paper maps. They only provide one of the many coverages used within an information system and it would be necessary digitally to overlay relevant coverages from other sources. Although the mathematics of intersecting lines and forming polygons and primitive regions is now tractable (ESRI, 1985; Wade et al, 1986), overlay analysis is computationally demanding. More importantly, paper-based overlay analysis involves judgement, since it takes into account a number of factors, including the accuracy and reliability of different data, a knowledge of the nature of underlying phenomena and of the purpose of the analysis. In general, automatic overlay analysis does not at present provide scope for the inclusion of such semantic knowledge and judgement, for example when resolving spurious polygons. This is clearly a topic for further research and development.

The lack of structured boundary data effectively means that the process of overlay analysis, which results in a catalogue of primitive regions, has to remain manual.

2.5 Homogeneity

Land parcels are defined to be homogeneous with respect to some critical attribute. Primitive regions are defined to be homogeneous with respect to a set of one or more attributes. BPUs should also contain populations with similar characteristics. But, the characteristic which must remain homogeneous will vary from application to application. Thus, no set of BPUs can satisfy this criteria adequately and this explains to some extent the plethora of data collecting, functional and reporting units in current use. Since spatial tessellations result in arbitrary cells, there is no scope for ensuring or engineering homogeneity. Such neutral units present some advantages in statistical processing (Visvalingam, 1983).

2.6 Modifiability

The problem of modifiable areal units is described in some detail elsewhere (DoE, 1987, p 165). This problem occurs when the requirement for homogeneity with respect to some application is violated either deliberately or unintentionally. It is well known that the design of electoral units can influence the electoral results and that different sets of areal units produce different results.

There is nothing intrinsically wrong with the use of optimising algorithms for the design of BSUs. Indeed, this is an essential requirement for the identification of functional units. All designed units impose some degree of bias to the data to facilitate meaningful analysis or use.

The problem occurs when such data are subsequently used for some other purpose, without regard to its inherent bias. Also, all BPUs, whose design involves the optimisation of some critical characteristic and which results in idiographic BSUs, are open to covert manipulation. Since tessellations result from the application of geometric rather than statistical criteria, they form arbitrary but neutral units which cannot be manipulated or modified easily.

2.7 Coverage

Land parcels and tessellations can provide a total coverage of space and populations. Unit postcodes do not provide a total coverage of space since they are collections of addresses. Many personal/population data do not relate exclusively to postal addresses. For example, a crime survey can relate information on offenders and victims to their respective postcodes. But, it is not clear as to which unit the data on the scene of the crime itself will be related if the incident occurred in a public park or a pathway between unit postcodes.

2.8 Size

Unless BPUs are aggregated with respect to customer-specified criteria, BPUs should be as small as possible so that they minimise heterogeneous groupings and provide maximum aggregational flexibility, without providing loopholes for the violation of the confidentiality of personal information.

Using these criteria, land parcels and addresses are too small for reporting purposes. If all land parcels and street addresses are Grid referenced, socio-economic data could be released for any set of spatially-defined BSUs. The latter include units with idiographic and regular shapes. The problem with the uniform grid is that a large number of cells located in rural areas tend to contain very small populations, leading to a suppression of data, whilst cells in urban areas contain large populations, leading to generalisations that are of limited value. Thus, if tessellations are to satisfy the size criterion, they must be data adaptive; i.e. in rural areas they must be large enough to preserve confidentiality and in urban areas they must be small enough to preserve homogeneity. This can be achieved through the use of nested tessellations. In the case of a grid framework, this effectively means that grid cells can be progressively subdivided within a consistent grid framework.

2.9 Ecological fallacy

Ecological fallacy refers to the unjustified inference of attributes about individuals from statistical generalisations about BPUs, i.e. groups of people. For example, the statistics may indicate that BPUs with relatively high proportions of immigrants also record above average levels of crime. It would be an error to conclude from this that immigrants are the source of crime, particularly if both immigrants and criminals are minority groups within these areas.

Openshaw (DoE, 1987, p 166) pointed out that designers of ecological classifications may label areas according to the relative concentration of target groups, which may form a very small minority of all residents. Such labels lead to misconceptions about an area's population profile or characteristic. Both examples show that the problems of ecological fallacy are basically problems of misrepresentation or misinterpretation of the results of statistical analyses and classifications by naive persons.

The scope for such misinterpretations can be reduced by the use of small homogeneous BPUs where possible.

2.10 Variability of base populations within BPUs

For statistical processing it is important that BPUs are designed to be near equal population units. Regular tessellations produce large variations in sample populations but nested tessellations may be data adaptive. Even irregular units, such as census EDs and unit postcodes, vary in size. Since land parcels are usually used to record nominal data, rather than statistical counts, the problem of variability in base populations does not occur with such coded data.

Variations in base populations produce a ratio bias. In area based analysis, the relative density of occurrence of phenomena is an important factor (Visvalingam, 1983a). Regular and nested tessellations provide some scope for accommodating this ratio bias since the area of BSUs is implicit. With irregular units, it is more difficult to deal with the ratio bias even when data on the area of BSUs are available. The spatial extent of unit postcodes is not defined. For the same reason, spatial surrogates alone are insufficient for area-based analysis.

2.11 Aggregational flexibility

BPUs may be conveniently aggregated using spatial centroids or nominal references in the form of hierarchic codenames. However, such aggregations are only meaningful if the BPUs nest within the units at higher levels.

Since the land parcels of street addresses are irregular, tessellations may be regarded as artificial agglomerations of such primitives. However, it is perfectly reasonable to view the aggregational process as the assignment of addresses to those cells which contain the bulk of that property. The use of visual centroids achieves this. The resulting errors are unlikely to prejudice statistical patterns to any significant extent. These errors have to be seen against the undefined boundaries of postcodes and the deliberate addition of random numbers to census counts to preserve confidentiality of data.

Postcodes do not always nest within Local Authority Districts at present. When BPUs do not nest neatly within application specified units, the counts have to be disaggregated and reaggregated to the required units. In the past, a grid tessellation has been used as a framework for disaggregation.

Overlay analysis forms the counterpart of disaggregation within the context of land information systems. Reaggregation, here, takes the form of a dissolution of boundaries between primitive regions which have the required set of attributes. Overlay analysis cannot be avoided altogether. But, it is more efficient to hold and process data for an integrated layer of primitive regions, compared with the repeated processing of separate coverages. General purpose GIS software, such as ARC/INFO (ESRI, 1985), allow both approaches. Although land parcels correspond to user-perceived BSUs, primitive regions will in time become the operational BSUs within computerised spatial information systems.

2.12 Stability

All irregular units are susceptible to change. The requirements for managing changes in data within an information system may be different. Many land information systems, such as planning systems, need to keep a record of changes to BSUs for purposes, such as land searches. They need continually to monitor and maintain a chronological record of change.

In contrast, the statistical analysis of snapshots of population characteristics requires comparable BPUs. There was thus a need to construct some slightly larger units to which both the 1971 and 1981 population census EDs could be aggregated in order to study change (DoE, 1987, p 164). Approximately 18,000 changes are made at the unit postcode level in Great Britain each year (DoE, 1987, p 172). The Postcode Address

File (PAF), which lists the addresses within postcodes, is updated every three months and is known to include some inconsistencies.

No idiographic unit, whether defined in spatial or aspatial terms, can remain constant. The retention of a constant set of idiographic units would be as inappropriate as the retention of some past standard industrial classification. Tessellations provide a stable framework for the statistical analysis of change, since they provide comparable, even if not constant, units and because they can provide a 100 percent coverage of space without incurring tremendous overheads for maintaining redundant units, which may be unused at present.

2.13 Historical and current usage

Land parcels have been used widely in many land and property based information systems. All human activity and building work must respect the rights to land and the restrictions upon its use (Dale, 1988). Land parcels formed the basis of the 1972 GISP proposals and of the Local Authority Management Information System (LAMIS) pioneered by Leeds, ICL and the DTI. The Tyne and Wear Area JIS survives because the economic benefits of having a property register is said to have convinced local politicians to continue and expand the system (DoE, 1987, p 167). The computerisation of the activities of Her Majesty's Land Registry (HMLR) and of others, such as the utilities and planning agencies, would stimulate the drive towards a systematic definition and use of land parcels.

Postcodes are being proposed as BSUs in place of EDs for reporting personal data (Chorley recommendation 41). Postcodes have been most widely used by commercial agencies for credit scoring and for generating customer/client profiles. It is becoming popular as a data collecting framework since all addresses can be tagged with a postcode and because more and more people are tending to remember and use their postcodes. The case for postcodes rests almost entirely on convenience of use and the scope that this offers for linking different sets of personal data.

Grid tessellations have been widely used for collecting, monitoring and mapping many types of spatial phenomena because they provide a stable framework for studying change and a convenient framework for mapping. The

1971 census provided statistics for 100 metre and 1 kilometre squares. The one-kilometre grid square data provided an acceptable framework for studying national, regional and more detailed patterns (C.R.U./O.P.C.S. /G.R.O., 1980) even if it was insufficient for studying detailed variations within neighbourhoods. These high resolution data also revealed the impact of the ratio bias in statistics for small areas, which remain concealed in many analyses of data for irregular areas (Visvalingam, 1978).

The one-kilometre grid square framework is also used by the Institute of Terrestrial Ecology for identifying land classes (Vincent, 1987). The USA Landsat, French SPOT, and other meteorological and microwave satellites generate rectangular or grid square data of various resolutions (Curran, 1985) which are extremely valuable for monitoring and predicting global and even local environmental phenomena. Tessellations of other shapes, for example irregular triangulated networks (Monmoniet, 1982), are also used for modelling and analysing continuous phenomena, such as terrain. Tessellations are particularly useful for modelling phenomena with gradual, rather than distinct and unambiguous, boundaries. The Military Survey is committed to producing digital terrain models (DTMs) for a 50 metre grid, based on contours on the Ordnance Survey 1:50 000 maps. The Institute of Hydrology is using such tessellated DTMs for generating fully connected and flow-directed networks of river systems and for defining catchment boundaries (Moore et al, 1985).

2.14 Storage, analysis and mapping using computers

If we ignore conceptual issues and consider tasks within a computerised information system, the choice of BSUs is largely one of choosing between vector and raster representations. The vector/raster debate is as complex as the consideration of BSUs and deserves separate systematic treatment. This debate is of relevance to applications which need to capture, store and process the boundaries of BSUs.

In general, the modelling and processing of vector data is much more complex than the storage and manipulation of tessellated data. Tessellations make overlay analysis of categorised data a trivial task. Categorised raster data can be compressed greatly using a linear quadtree data structure (Gargantini, 1982; Samet, 1984). Also, some form of implicit even if not explicit spatial tessellation is used to cluster and index data within vector systems in order to increase the efficiency of spatial search. But, the design of future information systems should not be unduly influenced by computational convenience.

3. DISCUSSION AND CONCLUSION

The choice of BSUs is an important decision for all GIS, whether manual or digital. We need to identify the most important factors, governing the choice of BSUs within different classes of applications. The following questions appear to be pertinent.

Which is the largest scale of interest to your application?

- Is it global, national, regional, local or even more detailed?

- What impact has the range of scales on requirements for accuracy, size, shape, coverage and aggregational flexibility of BSUs?

- How does scale influence the choice of the critical attributes, which must remain homogenous within BSUs at different levels?

What are the goals and tasks within your application?

- Is it largely observation for monitoring (e.g. in some applications of satellite data), recording for management and operations (e.g. in facilities management), monitoring for control (e.g. planning), comparison for action (e.g. new development), or analysis for targeting (e.g. services)?

What type of entity does your application focus upon and what is its nature?

- Is it space and land (ownership, properties, content, constraints on use), statistical populations (people, livestock, energy levels etc. and their average and exceptional conditions) or events (all or some)?

- Does it have well-defined boundaries or does it vary continuously in space? Is the characteristic of interest predefined (e.g. administrative boundaries), directly observable and/or verifiable (e.g. walls and heights or intervisibility) or has it to be inferred from proxy measures (e.g. deprivation and markets)?

Different answers to these questions will place different interpretations and emphasis on the issues relating to BSUs and may identify conflicting requirements. The choice of BSUs will also have to be moderated by pragmatic considerations, such as state-of-the-art limitations on the collection, storage and processing of information (e.g. satellite imagery offering at best 10 metre resolution; the speed of capture and structure of OS digital data), legal constraints (e.g. the Data Protection Act), or clerical and administrative convenience (e.g. the unit postcode forming a part of the postal address).

LIS tend to be large-scale applications with a requirement for continual monitoring of accurately defined boundaries, especially of legal and financial land parcels. The boundaries are either predefined or tangible and form meaningful lines of demarcation.

However, many spatial phenomena do not have well-defined boundaries. When phenomena change gradually from one type to another, boundaries become fuzzy and arbitrary (Burrough, 1986) and a high resolution spatial tessellation may be more appropriate for many environmental monitoring systems. Data for tessellations allow users to impose their own interpretations and classify space as appropriate for their applications.

Populations, similarly, are not classified according to application-specific types by data collecting agencies. Instead, a range of descriptive statistics are released for each BPU. Applications classify these populations and their neighbourhoods, using a selection of data and appropriate techniques, although an increasing number of users find it more convenient to work with a proprietary classification, such as ACORN. Such classifications may be based on data for either high resolution tessellations or irregular units. But, since addresses have not been grid referenced as yet, census data have been mainly released for irregular units. The Chorley Committee has recommended that unit postcodes be adopted as a standard data reporting framework so as to avoid assumption-based disaggregation of spatial statistics when linking data from different sources.

But, will unit postcodes satisfy the requirements of future corporate GIS which may need to relate legal, land, population, environmental and other data? Unit postcodes have known deficiencies; they are BPUs and not BSUs; they vary with respect to base populations; they do not provide a stable framework for comparing statistical snapshots; they do not nest neatly within reporting units such as Local Authority districts; they do not cover all space of interest; and, there is some concern that postcodes may be used to work around the 1984 Data Protection Act.

Commercial agencies have tended to use stereotypes for quick identification of target groups. Socio-economic class was used initially but neighbourhood profiles, based on population census and market research data, have become more popular in the 1980s. Stereotypes serve to increase the efficiency, rather than the equity, of targeting (Visvalingam, 1983b). Unit postcodes provide a convenient framework for interrelating proxy data and for targetting, especially through direct mail. Recently, the Data Protection Registrar has objected to the inequitable practice of using an address as a cheap and easy predictor of credit worthiness (Daily Telegraph, p 17). Note that the use of generalised stereotypes in area-based decision-making is even less discriminating.

To be fair, many decisions are targeted at groups, rather than at individuals. Unit postcodes will have to be redesigned and spatially defined to act as BSUs. Will they then correspond to the postman's walk; i.e. should we still regard them as postcodes? Or, will they in effect become some other spatial reference to be attached to an address? If the proposal to Grid reference street addresses is implemented, will postcodes remain as attractive given that a spatial reference offers greater aggregational flexibility than a nominal reference?

The Chorley Committee was asked to advise on the future handling of geographic information taking into account modern developments in IT. The case for postcodes rests on the fact that they are widely known and used. This was convenient for data collection and processing in the past. But, information systems of the future should not have to rely on people remembering and correctly using nominal and/or cadastral addresses. Already, we are carrying more and more cards for one purpose or another and many of these are electronically processed even at present. How many of us know, let alone remember, the barcodes on personal documents? Optical information cards, also known as LaserCards, are already in use by the US government for US Army training and by health insurance companies for personal health cards. Thus, information systems of the future may not need the convenience of easily memorable postcodes.

Even if memorability remains an important requirement, this does not in itself establish the case for postcodes. Cadastral addresses and personal identity numbers are used widely in some European and Third World countries respectively. Also, most people have little difficulty in remembering and using their own and several other telephone numbers with 10 or more digits. The design of BSUs for a future national GIS must be based on substantive rather than pragmatic criteria.

The need for a variety of functional units will continue. OPCS/GRO(S), ESRC and ICL are currently providing funds to Birkbeck College for development of an experimental on-line service for querying a database of unaggregated census records. Users will be able to extract any aggregate information - for whatever area(s) or groups of people and cross-tabulated as required - unless this is likely to disclose details of an identifiable individual or household or lead to unreasonable intrusions into privacy (Rhind and Higgins, 1988). Given Grid referenced primitives, data could quite easily be provided for any set of spatially defined BSUs. If on-line services of this kind are acceptable to the public, then it is quite likely that much government data, if released, will be disseminated in this way for user-defined units where possible. Within this scenario, there may well be less interest in standard BSUs.

To conclude, unit postcodes as we know them now have several disadvantages. For some applications, they offer administrative and mailing convenience. But, no set of irregular functional BSUs can act as standard units for any length of time. Only tessellations can provide a stable framework for analysing change, for comparative studies and for modeling continuous phenomena, such as terrain. However, many applications will continue to use a variety of irregular, functional BSUs for substantive reasons.

ACKNOWLEDGMENTS

I am grateful for the opinions expressed within the context of a debate on BSUs by my colleague, Dr R N Davidson, and by Geography and Computer Science students of the Final Year Computer Cartography option but I assume full responsibility for the views expressed in this paper. I am also grateful to Drs Graham Kirby and Mike Turner of the CISRG for their comments on this paper.

REFERENCES CITED

Burrough, P. A. (1986) **Principles of Geographical Information Systems for Land Resources Assessment,** Clarendon Press, Oxford.

C.R.U./O.P.C.S./G.R.O. (1980) **People in Britain - a census atlas,** H.M.S.O., London.

Curran, P. J. (1985) **Principles of Remote Sensing,** Longman.

Daily Telegraph (1988) "Credit checks get October ultimatum", Saturday 21st May 1988, p 17.

Dale, P. F. (1988) "Land Information Systems and the land parcel", **Mapping Awareness** 2 (2), 27 - 29.

Department of the Environment (1987) **Handling Geographic Information,** Report of the Committee of Enquiry chaired by Lord Chorley, H.M.S.O., London.

ESRI (1985) **ARC/INFO Users Manual - Version 3,** Environmental Systems Research Institute, Redlands, California.

Gargantini, I. (1982) "An effective way to represent quadtrees" **Comm ACM** 25 (12), 905 - 910.

Kirby, G. H., Visvalingam, M. and Wade, P. (1989), "The recognition and representation of polygons with holes", **Computer Journal,** in press.

Mapping Awareness (1988), "The national Mapping Awareness conference", **Mapping Awareness** 1 (10), Feb. 1988, 2 - 5.

Monmonier, M. (1982) **Computer-Assisted Cartography : principles and prospects,** Prentice-Hall.

Moore, R. V., Morris, D. G. and Venn, M. W. (1984) "Towards a CODASYL database for the UK river network", paper presented at the Int. Symp. on Spatial Data Handling, August 1984, Zurich, Switzerland.

Peuquet, D. J. (1984) "A conceptual framework and comparison of spatial data models", **Cartographica** 21 (4), 66 - 113.

Rhind, D. and Higgins, M. (1988) "Customer-selected results from the 1991 Census", **BURISA** 82, 7 - 9.

Samet, H. (1984) "The quadtree and related hierarchical data structures", **ACM Computing Surveys,** June 1984, 187-260.

Vincent, P. (1987) "Geographical Information Systems at Lancaster", BURISA 80, Sept. 1987, p 9.

Visvalingam, M. (1978) "The signed chi-square measure for mapping", **Cartographic J.** 15 (2), 93 - 98.

Visvalingam, M. (1983a) "The operational definition of area-based social indicators", **Environment and Planning,** A 15, 831 - 839.

Visvalingam, M. (1983b) "An examination of some criticisms of area-based policies", **Census Research Unit Working Paper** 22, Department of Geography, University of Durham, 25 pp.

Visvalingam, M. and Kirby, G. H. (1984) "The impact of advances in Information Technology on the cartographic interface in social planning", **Miscellaneous Series** 27, Department of Geography, University of Hull.

Visvalingam, M., Wade, P. and Kirby, G.H. (1986) "Extraction of area topology from line geometry", Blakemore, M. (ed.) **Auto Carto London** volume 1, 156 - 165.

Wade, P., Visvalingam, M. and Kirby, G.H. (1986) "From line geometry to area topology", **Cartographic Information Systems Research Group Discussion Paper 1**, University of Hull.